Estimation and empirical performance of non-scalar dynamic conditional correlation models

Luc Bauwens[1], Lyudmila Grigoryeva[2], and Juan-Pablo Ortega[3]

Abstract

A method capable of estimating richly parametrized versions of the dynamic conditional correlation (DCC) model that go beyond the standard scalar case is presented. The algorithm is based on the maximization of a Gaussian quasi-likelihood using a Bregman-proximal trust-region method that handles the various non-linear stationarity and positivity constraints that arise in this context. The general matrix Hadamard DCC model with full rank, rank equal to two and, additionally, two different rank one matrix specifications are considered. In the last mentioned case, the elements of the vectors that determine the rank one parameter matrices are either arbitrary or parsimoniously defined using the Almon lag function. Actual stock returns data in dimensions up to thirty are used in order to carry out performance comparisons according to several in- and out-of-sample criteria. Empirical results show that the use of richly parametrized models adds value with respect to the conventional scalar case.

Keywords: Multivariate volatility modeling, Dynamic conditional correlations (DCC), Non-scalar DCC models, Constrained optimization, Bregman divergences, Bregman-proximal trust-region method.

JEL Classification: C13, C32, G17.

[1]Université catholique de Louvain, CORE, Voie du Roman Pays 34 L1.03.01, B-1348 Louvain-La-Neuve. Belgium; University of Johannesburg, Department of Economics, Johannesburg, South Africa. luc.bauwens@uclouvain.be

[2]Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, UFR des Sciences et Techniques. 16, route de Gray. F-25030 Besançon cedex. France. Lyudmyla.Grygoryeva@univ-fcomte.fr

[3]Centre National de la Recherche Scientifique, Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, UFR des Sciences et Techniques. 16, route de Gray. F-25030 Besançon cedex. France. Juan-Pablo.Ortega@univ-fcomte.fr

Acknowledgments: Lyudmila Grigoryeva and Juan-Pablo Ortega acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493). Lyudmila Grigoryeva acknowledges financial support from the Faculty for the Future Program of the Schlumberger Foundation. Luc Bauwens acknowledges the support of "Projet d'Actions de Recherche Concertées" 12/17-045 of the "Communauté francaise de Belgique", granted by the "Académie universitaire Louvain".

1 Introduction

The choice of the dynamic conditional correlation (DCC) model has become very common in the multivariate GARCH applied literature, where the goal is often to fit the dynamics of time-varying conditional variances and correlations of asset returns and to forecast their future values. The DCC family was developed in [Eng02] and [TT02] as a generalization of the constant conditional correlation (CCC) model of [Bol90]. The main advantage of DCC models is the availability of a two-step estimation procedure which, combined with correlation targeting, makes their use feasible even when the number of assets is high. It is worth noting that even though in the original paper [Eng02] a general matrix Hadamard-type model parameterization is proposed, it is almost exclusively the scalar prescription that is used in applications. This simplified version of the model imposes the same correlation dynamics to all the pairs of assets that are considered which, for sizable dimensions, may constitute an excessively restrictive homogeneity assumption.

The main contribution of this paper is to provide adequate estimation tools for several non-scalar richly parameterized DCC models and to empirically evaluate how they perform with respect to each other and to the scalar model. The non-scalar models that are considered are the general matrix Hadamard-type model and four more parsimonious particular cases thereof. The main conclusions are that the estimation of these models is practically feasible in moderate dimensions (up to thirty), and that two of the non-scalar models considered are worth using in practice.

This study is restricted to DCC models in which only an approximate correlation targeting as suggested by [Eng02] is possible. This is a widespread approach that aims at reducing the number of parameters in the likelihood maximization by replacing the constant term matrix of the quasi-correlation process by a moment estimate. This approach is controversial in the DCC context since the approximate targeting procedure is statistically inconsistent (see [CM12] for an extensive review). This issue has motivated the introduction by [Aie13] of a corrected DCC (cDCC) model that carefully addresses it, but the price to pay is that estimation is more convoluted. Despite the substantial theoretical interest of the cDCC model, it is not considered in this work since existing empirical findings and simulation results in [Aie13] reveal that the performance of the cDCC and DCC models does not differ much in practice for typical financial series.

The need to go beyond the scalar DCC prescription has been partly addressed in [HF09] and [CES06]. In these papers, a so-called diagonal DCC model that allows for asset-specific heterogeneity in the correlation structure is used. This extension specifies the parameter matrix associated to the lagged innovations as a rank one matrix, and the same applies to the parameter matrix of the lagged quasi-correlation term although that one is also kept as a scalar parameter in [HF09]. The results in those papers provide empirical evidence supporting that these richer DCC models exhibit improved performance with respect to the scalar model. Another relevant paper in the same direction is that of [NSS14] who introduce (among others) a model, called Rotated DCC (RDCC), which uses similar specifications as in the two previously cited papers, after applying an estimated orthogonal transformation to the devolatized returns, and illustrate that it yields increased performance when compared to existing models like, for instance, the OGARCH model of [Ale98]. The contribution of this paper with respect to these mentioned works is that other non-scalar and more richly parametrized models are proposed and estimated.

The results in [HF09] and [CES06] lead us to believe that non-scalar DCC models with asset specific dynamics in the correlation structure, like the one associated to the original matrix Hadamard-type parameterization of [Eng02], can yield a superior performance in practical applications when compared with the widespread scalar DCC model. This conjecture can only be verified when effective estimation procedures are available for the richer models, the absence of which explains in part statements in the literature (see for example [BCG06] or Chapter 7 of [Eng09]) about the lack of empirical interest of these more general models. Two difficulties in this respect arise: the first one is the quadratic dependence of

the number of parameters on the model dimension and the second is the need to impose at the time of estimation the nonlinear constraints that ensure the positivity and the stationarity of the dynamic conditional correlation process.

Our attention hence turns first to the optimization techniques that can be used for the estimation of non-scalar DCC models and, more specifically, to the approach proposed by [CO14] based on the use of Bregman divergences, that we take as our starting point. That paper successfully applies this approach to the estimation of the heavily parametrized VEC-GARCH model subjected to stationarity and positivity constraints. We extend this optimization method to the DCC models and develop explicit estimation tools for a variety of non-scalar DCC specifications originating from the general Hadamard-type DCC prescription in [Eng02]. Even though the DCC family has a much smaller parameter space than VEC, the use of the Bregman divergences approach is extremely advantageous in the treatment of the DCC highly nonlinear optimization constraints. The paper is therefore organized around the three topics that we describe in the following paragraphs.

DCC model specifications: The considered models are the Hadamard DCC family for which the two parameter matrices of the lagged innovations and lagged quasi-correlation terms are symmetric with full rank, as well as four other subfamilies where these matrices have smaller ranks. More specifically, in one case the rank is set equal to two, and in three other cases it is equal to one. The first of these three "rank one" cases is equivalent to the diagonal DCC model considered also in [HF09] and [CES06], where the parameter matrices are built as outer products of vectors of size the number n of assets that are being modeled. The other two "rank one" models are new and called Almon DCC and Almon shuffle DCC. In these models, the elements of the vectors that generate the rank deficient parameter matrices are defined using an Almon function (see [Alm65]). Thus, like in the scalar model, the number of parameters that need to be estimated does not depend on the dimension n but the correlation dynamics differs for all pairs of assets. The Almon DCC models are therefore more flexible than the scalar DCC model, while determined by a comparable number of parameters. Section 2 is devoted to presenting the general setup for DCC models, to describing in detail the different parameterizations under study, as well as the constraints that are imposed on each of them in order to ensure the stationarity of the process and the positive definiteness of the resulting conditional correlation matrices.

DCC model estimation: This is the subject of Section 3, which presents the main ideas behind the optimization algorithm chosen for the estimation. Bregman divergences are used in order to handle the model constraints, following the scheme proposed in [CO14]. This approach is very popular in the context of machine learning (see for instance [DT07] and [KSD09a]). In our situation, it is particularly advantageous because it allows for the treatment of the nonlinear optimization constraints that the problem is exposed to without resorting to Lagrange duality or other techniques that demand the solution of supplementary optimization problems. Section 3 contains a comprehensive description of the ingredients necessary to implement this optimization algorithm for each of the DCC models.

DCC specifications performance assessment: The possibility of estimating non-scalar DCC models with the tools just mentioned, allows us to empirically study their performance and, ultimately, to assess the need for those models in the processing of financial data. In Section 4 an in- and out-of-sample study is carried out using a dataset containing the returns of the thirty assets listed in the Dow Jones Industrial Average Index. The results reveal that: **(i)** no model (scalar or not) dominates systematically the others in terms of in-sample and out-of-sample performance; **(ii)** the full rank Hadamard model is in most cases dominated by less richly parameterized choices; **(iii)** among the more parsimonious models, the rank one (or diagonal) model and the Almon (shuffle) models perform better in many cases than the scalar model. Nevertheless, for a given dataset, the practical choice between the models has to be made by taking into account the specifics of the time series in question, the available sample length, dimension, and other considerations.

A technical appendix is available online. It contains the proofs of the results in the paper and many other technical details that could not be included in the main body of the paper due to a space

limitations.

2 Dynamic conditional correlation models

The different dynamic conditional correlation (DCC) model specifications studied in the paper are presented in this section. In the first subsection, the most general model is defined; it is referred to as the (general) Hadamard DCC model. The following sections define more particular models obtained by imposing parametric restrictions on the general Hadamard case. For each model, the parametric constraints that need to be imposed so that the correlation process admits a stationary solution and the resulting conditional correlation matrices are positive definite, are explained. In some cases, identification constraints are also required.

All DCC models are based on the following functional prescription for the n-dimensional conditionally heteroscedastic discrete-time process $\{\mathbf{r}_t\}$:

$$\mathbf{r}_t = H_t^{1/2}\boldsymbol{\xi}_t, \ \ \{\boldsymbol{\xi}_t\} \sim \mathrm{IN}(\mathbf{0}_n, \mathbb{I}_n), \quad t = 1, 2, \ldots, T, \tag{2.1}$$

meaning that $\{\boldsymbol{\xi}_t\}$ is a set of n-dimensional independent normally distributed random vectors with mean $\mathbf{0}_n$ and identity covariance matrix $\mathbb{I}_n$. $\{H_t\}$ is a predictable positive semidefinite matrix process, that is, for each $t \in \mathbb{N}$, H_t is a random matrix that takes values in $\mathbb{S}_n^+$ (the cone of positive semidefinite symmetric matrices); predictable means that the random variable H_t is $\mathcal{F}_{t-1}$-measurable, where $\mathcal{F}_{t-1} := \sigma(\mathbf{r}_1, \ldots, \mathbf{r}_{t-1})$ is the information set generated by $\{\mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\}$.

In the first stage of the DCC model construction, a dynamic process is chosen for the conditional variances $\sigma_{i,t}^2$ of each component $r_{i,t}$ of $\mathbf{r}_t$, for example the GARCH(1,1) model of [Bol86]:

$$\sigma_{i,t}^2 = \alpha_{0,i} + \alpha_{1,i} r_{i,t-1}^2 + \beta_{1,i}\sigma_{i,t-1}^2, \quad i \in \{1, \ldots, n\}, \tag{2.2}$$

where the parameters $\alpha_{0,i}$, $\alpha_{1,i}$, and $\beta_{1,i}$ for all $i \in \{1, \ldots, n\}$ satisfy the inequalities $\alpha_{0,i} > 0$, $\alpha_{1,i}, \beta_{1,i} \geq 0$, and $\alpha_{1,i} + \beta_{1,i} < 1$, in order to ensure the stationarity of the process and the positivity of the conditional variances $\sigma_{i,t}^2$. Standardized returns of the i-th asset at time t are defined as $\varepsilon_{i,t} := r_{i,t}/\sigma_{i,t}$ for $i = 1, 2, \ldots, n$ and are assembled in the vector $\boldsymbol{\varepsilon}_t$. More sophisticated GARCH models could be used in (2.2) to take into account specific stylized facts of the financial time series at hand like, for example, asymmetry.

The second stage of the DCC model construction consists in specifying a dynamic equation for the conditional correlation matrices R_t of the standardized returns $\boldsymbol{\varepsilon}_t$. The matrix R_t is related to the covariance matrix H_t by the relation

$$H_t = D_t R_t D_t, \tag{2.3}$$

where $D_t := \mathrm{diag}(\sigma_{1,t}, \ldots, \sigma_{n,t})$. The dynamic behavior of the conditional correlation process $\{R_t\}$ is modeled through a dynamic matrix process $\{Q_t\}$ to which it is connected by the relation

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \tag{2.4}$$

where the matrix Q_t^* is defined as $Q_t^* := \mathrm{Diag}(Q_t)$ (the Diag operator sets equal to zero all the components of a square matrix except for those that are on the main diagonal). The relation (2.4) transforms Q_t into a correlation matrix since the diagonal elements of Q_t are not necessarily equal to 1. The different DCC models differ in the way in which the dynamics of Q_t is parameterized.

2.1 The general DCC model and its parameter constraints

We consider separately the targeted and the non-targeted versions of the general DCC model because they require different parameter constraints.

Non-targeted general DCC model. The most general dynamic equation used for Q_t is (see [Eng02]):

$$Q_t = C + A \odot (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top) + B \odot Q_{t-1}, \tag{2.5}$$

where A, B, and C are symmetric parameter matrices, that is, $A, B, C \in \mathbb{S}_n$. The functional form of the equation (2.5) guarantees neither that the resulting joint process $\{\mathbf{r}_t, H_t\}$ is stationary nor that $\{H_t\}$ takes values in $\mathbb{S}_n^+$ and hence consists of covariance matrices. Sufficient conditions on the parameters A, B, and C that ensure those features are:

$$\textbf{(SC) Stationarity constraints:} \qquad |A_{ij} + B_{ij}| < 1, \quad i,j \in \{1,\dots,n\}. \tag{2.6}$$

$$\textbf{(PC) Positivity constraints:} \qquad A \succeq 0, \quad B \succeq 0, \ C \succ 0, \ \text{and} \ Q_0 \succeq 0. \tag{2.7}$$

In these expressions the symbols $\succeq 0$ (respectively, $\succ 0$) denote positive semidefiniteness (respectively, positive definiteness). We emphasize that there is still no proof in the DCC context that ensures that what we called the stationarity constraints (2.6) actually guarantee that dynamical feature. Such proof is only available for the cDCC model of [Aie13]. The sufficiency of the positivity constraints follows from an inductive argument using the expression (2.5) that defines $\{Q_t\}$, together with the Schur Product Theorem (see [BR97]).

The specification (2.5) is rarely used in practice due to the large number of parameters involved which makes this version of the DCC model suffer from the so called curse of dimensionality. For that reason, in the sequel, we focus mainly on DCC models with approximate correlation targeting, even though we also provide specific remarks concerning the non-targeted case (2.5).

Targeted general DCC model. The targeted version of the general DCC model is obtained from (2.5) by replacing the constant term C by $(\mathbf{i}_n\mathbf{i}_n^\top - A - B) \odot S$, namely,

$$Q_t = (\mathbf{i}_n\mathbf{i}_n^\top - A - B) \odot S + A \odot (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top) + B \odot Q_{t-1}, \tag{2.8}$$

where $\mathbf{i}_n$ stands for the vector of length n consisting of ones, $A, B \in \mathbb{S}_n$ are parameter matrices, and $S = E\left[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top\right] \in \mathbb{S}_n^+$ (see [CM12], and [Aie13] for additional details). This is motivated by the possibility to estimate S, prior to the estimation of A and B, by

$$\hat{S} := \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top. \tag{2.9}$$

Sufficient conditions on the parameters that ensure the stationarity of the joint process $\{\mathbf{r}_t, H_t\}$ and the positive definiteness of $\{H_t\}$ are:

$$\textbf{(PC) Positivity constraints:} \qquad A \succeq 0, \quad B \succeq 0, \tag{2.10}$$

$$\left(\mathbf{i}_n\mathbf{i}_n^\top - A - B\right) \odot S \succ 0, \quad Q_0 \succeq 0. \tag{2.11}$$

It is necessary to stress that the first constraint in (2.11) cannot be replaced by $\mathbf{i}_n\mathbf{i}_n^\top - A - B \succ 0$ even if $S \succ 0$ (which is the case when S is estimated using $\hat{S}$ in (2.9)). Indeed, if the constraints (2.10) hold, they generically prevent that $\mathbf{i}_n\mathbf{i}_n^\top - A - B \succ 0$ since any vector $\mathbf{v}$ which is simultaneously in the kernel of $\mathbf{i}_n\mathbf{i}_n^\top$ and not in the kernel of $G := A + B$ satisfies $\langle \mathbf{v}, (\mathbf{i}_n\mathbf{i}_n^\top - A - B)\mathbf{v}\rangle < 0$. In the targeted case there is no need to impose stationarity constraints due to the following result[4] that shows that the stationarity constraints are automatically satisfied in the presence of the positivity constraints (2.10)-(2.11) (see the technical appendix for the proof).

Proposition 2.1 *Consider the targeted DCC model (2.1), (2.3), (2.4), with Q_t as in (2.8). If the parameter matrices $A, B \in \mathbb{S}_n$ satisfy the positivity constraints (2.10) and (2.11), then the stationarity constraints (2.6) are automatically satisfied.*

[4] We thank an anonymous referee for this valuable observation.

In the following subsections we specify several parameter subfamilies of the general DCC model that we just introduced, together with their intrinsic parameterizations and associated constraints. Notice that in the targeted models setup, Proposition 2.1 remains valid for all the cases considered.

2.2 The Hadamard DCC model and its parameter constraints

Since the parameter matrices A and B are generic elements in $\mathbb{S}_n$, they can be parametrized with $(\boldsymbol{a},\boldsymbol{b}) \in \mathbb{R}^N \times \mathbb{R}^N$, where $N = \frac{1}{2}n\,(n+1)$, by setting $A := \operatorname{math}(\boldsymbol{a})$ and $B := \operatorname{math}(\boldsymbol{b})$, where the operator $\operatorname{math} : \mathbb{R}^N \longrightarrow \mathbb{S}_n$ is the inverse of the vech operator (see the technical appendix for the details). The space $\mathbb{R}^N \times \mathbb{R}^N$ is referred to as the intrinsic parameter space of the Hadamard DCC model and is generically denoted by $\mathcal{P} \times \mathcal{P}$. The dimension of the intrinsic parameter subspace $\mathcal{P}$ is denoted by P. Using these intrinsic parameters, the relation (2.8) is written as

$$Q_t = \left(\mathbf{i}_n\mathbf{i}_n^\top - \operatorname{math}(\boldsymbol{a}) - \operatorname{math}(\boldsymbol{b})\right) \odot S + \operatorname{math}(\boldsymbol{a}) \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top\right) + \operatorname{math}(\boldsymbol{b}) \odot Q_{t-1}. \tag{2.12}$$

Sufficient conditions on the parameters $\boldsymbol{a}$ and $\boldsymbol{b}$ that ensure that the resulting joint process $\{\mathbf{r}_t, H_t\}$ is stationary and that $\{H_t\}$ takes values in $\mathbb{S}_n^+$ and hence consists of covariance matrices are:

(PC) Positivity constraints:
$$\operatorname{math}(\boldsymbol{a}) \succeq 0, \quad \operatorname{math}(\boldsymbol{b}) \succeq 0, \text{ and} \tag{2.13}$$
$$\left(\mathbf{i}_n\mathbf{i}_n^\top - \operatorname{math}(\boldsymbol{a}) - \operatorname{math}(\boldsymbol{b})\right) \odot S \succ 0, \quad Q_0 \succeq 0. \tag{2.14}$$

In the non-targeted Hadamard setup, the parameter matrix $C \in \mathbb{S}_n$ in (2.5) is parametrized by $\boldsymbol{c} \in \mathbb{R}^N$, $N = \frac{1}{2}n\,(n+1)$ via the assignment $C := \operatorname{math}(\boldsymbol{c})$, and hence the constraints on the parameters $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ are:

(SC) Stationarity constraints:
$$|a_i + b_i| < 1, \quad i \in \{1,\dots,N\}. \tag{2.15}$$
(PC) Positivity constraints:
$$\operatorname{math}(\boldsymbol{a}) \succeq 0, \ \operatorname{math}(\boldsymbol{b}) \succeq 0, \ \operatorname{math}(\boldsymbol{c}) \succ 0, \ Q_0 \succeq 0. \tag{2.16}$$

Given that the number of parameters in (2.12) exhibits a quadratic dependence on the dimension n of the process $\{\mathbf{r}_t\}$, the following subsections are dedicated to the formulation of several more parsimonious subparameterizations of this general Hadamard DCC model and to the characterization of the associated parameter constraints.

2.3 Rank deficient DCC models

The rank deficient DCC models constitute a subfamily of the Hadamard DCC family where the matrices A and $B \in \mathbb{S}_n$ in (2.8) are constrained to have a common prescribed rank $r \in \{1,\dots,n-1\}$, that is, $\operatorname{rank}(A) = \operatorname{rank}(B) = r$. A natural parameterization for A and B consists of using $\widetilde{A}, \widetilde{B} \in \mathbb{M}_{n,r}$ (the space of real $n \times r$ matrices) such that $A = \widetilde{A}\widetilde{A}^\top$ and $B = \widetilde{B}\widetilde{B}^\top$. This choice poses an identification problem due to the invariance properties of the product $\widetilde{A}\widetilde{A}^\top$. Indeed, given $A \in \mathbb{S}_n$ with rank r and $\widetilde{A} \in \mathbb{M}_{n,r}$ such that $A = \widetilde{A}\widetilde{A}^\top$, this equality also holds true for any other matrix $\tilde{\widetilde{A}} := \widetilde{A}O$, with $O \in \mathbb{O}(r)$ an arbitrary element of the orthogonal group in r dimensions: $\tilde{\widetilde{A}}\tilde{\widetilde{A}}^\top = \widetilde{A}OO^\top\widetilde{A}^\top = \widetilde{A}\widetilde{A}^\top = A$. This observation indicates that the intrinsic parameter subspace looked for is not $\mathbb{M}_{n,r}$ but the orbit space $\mathbb{M}_{n,r}/\mathbb{O}(r)$ of the Lie group action of $\mathbb{O}(r)$ on the set of rectangular matrices $\mathbb{M}_{n,r}$ via the map $\Phi : \mathbb{O}(r) \times \mathbb{M}_{n,r} \longrightarrow \mathbb{M}_{n,r}$ given by $(O, A) \longmapsto AO^{-1}$. The next proposition provides a convenient model space for the quotient $\mathbb{M}_{n,r}/\mathbb{O}(r)$ and, as a corollary, a characterization of the intrinsic parameter subspace of a rank deficient DCC model. A proof of this result can be found in Section C of the TA.

Proposition 2.2 *In the setup just described, consider the cone $\mathbb{L}_{n,m}^+$ of lower triangular matrices with positive elements in the main diagonal. Then, the map $\Psi : \mathbb{L}_{n,m}^+ \longrightarrow \mathbb{M}_{n,r}/\mathbb{O}(r)$ given by $A \longmapsto [A]$ is*

a bijection. The symbol $[A]$ *denotes the orbit in* $\mathbb{M}_{n,r}/\mathbb{O}(r)$ *corresponding to the element* $A \in \mathbb{M}_{n,r}$ *with respect to the action of* $\mathbb{O}(r)$ *on* $\mathbb{M}_{n,r}$.

This proposition implies that the intrinsic parameter subspace of the rank deficient family is $\mathbb{L}^{+}_{n,m}$. This can be described as $\mathbb{R}^{N^*}$, $N^* = nr - \frac{1}{2}r(r-1)$, via the operator $\mathrm{mat}_r : \mathbb{R}^{N^*} \longrightarrow \mathbb{L}_{n,r}$ that transforms a vector of length N^* into the lower triangular $n \times r$ matrix defined by,

$$\mathrm{mat}_r(\mathbf{v}) = \begin{pmatrix} v_1 & 0 & \cdots & 0 \\ v_2 & v_{n+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ v_r & v_{n+r-1} & \cdots & v_{N^*-n+r} \\ \vdots & \vdots & \ddots & \vdots \\ v_n & v_{2n-1} & \cdots & v_{N^*} \end{pmatrix}, \quad \text{for any } \mathbf{v} \in \mathbb{R}^{N^*}, \tag{2.17}$$

just by adding a positivity constraint on the entries of the vector in $\mathbb{R}^{N^*}$ that constitute the main diagonal of the corresponding matrix via the mat_r operator. More explicitly, let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{N^*}$ and define $\widetilde{A} := \mathrm{mat}_r(\boldsymbol{a})$ and $\widetilde{B} := \mathrm{mat}_r(\boldsymbol{b})$. The matrices A and B in (2.8) are hence given by $A := \widetilde{A}\widetilde{A}^\top$, $B := \widetilde{B}\widetilde{B}^\top$. Using the notation just adopted, the rank deficient DCC model specification can be written as

$$Q_t = (\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{A}\widetilde{A}^\top - \widetilde{B}\widetilde{B}^\top) \odot S + (\widetilde{A}\widetilde{A}^\top) \odot (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top) + (\widetilde{B}\widetilde{B}^\top) \odot Q_{t-1}. \tag{2.18}$$

In order to ensure that $\widetilde{A}, \widetilde{B} \in \mathbb{L}^{+}_{n,r}$ and that, by Proposition 2.2, the model is well identified, the following constraints need to be imposed:

$$\textbf{(IC) Identification constraints:} \qquad a_{i_j} > 0, \quad b_{i_j} > 0, \tag{2.19}$$

where $i_j = n(j-1) + \frac{1}{2}j(3-j)$, $j \in \{1, \ldots, r\}$ are the entries of any vector in $\mathbb{R}^{N^*}$ that amount to the main diagonal of the corresponding matrix in $\mathbb{L}_{n,r}$ via the mat_r representation.
Additionally, the positive definiteness constraints are expressed as:

$$\textbf{(PC) Positivity constraints:} \qquad (\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{A}\widetilde{A}^\top - \widetilde{B}\widetilde{B}^\top) \odot S \succ 0, \quad Q_0 \succeq 0. \tag{2.20}$$

In the non-targeted version of the rank deficient DCC model, the additional parameter matrix $C \in \mathbb{S}_n$ in (2.5) is parametrized by $\boldsymbol{c} \in \mathbb{R}^N$, $N = \frac{1}{2}n(n+1)$ via the assignment $C := \mathrm{math}(\boldsymbol{c})$. In this case, the stationarity constraints need to be separately added; more specifically, we have:

$$\textbf{(IC) Identification constraints:} \qquad a_{i_j} > 0, \ b_{i_j} > 0, \tag{2.21}$$
$$\text{with } i_j = n(j-1) + \frac{1}{2}j(3-j), \ j \in \{1, \ldots, r\}.$$

$$\textbf{(PC) Positivity constraints:} \qquad \mathrm{math}(\boldsymbol{c}) \succ 0, \quad Q_0 \succeq 0. \tag{2.22}$$

$$\textbf{(SC) Stationarity constraints:} \qquad \sum_{k=1}^{r} \left| \widetilde{A}_{ik}\widetilde{A}_{jk} + \widetilde{B}_{ik}\widetilde{B}_{jk} \right| < 1, \ i, j \in \{1, \ldots, n\}, \ i \geq j. \tag{2.23}$$

2.4 The Almon DCC model

The Almon DCC model specification is a particular case of the rank deficient DCC model with $r = 1$, where the vectors that generate the matrices $A, B \in \mathbb{S}_n$ in (2.8) are parametrized using the Almon lag

function of [Alm65] alm_n defined as follows: given $n \in \mathbb{N}$ and $\mathbf{v} \in \mathbb{R}^3$, the Almon lag operator $\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$ generates a vector $\mathrm{alm}_n(\mathbf{v})$ whose entries are $(\mathrm{alm}_n(\mathbf{v}))_i = v_1 + \exp(v_2 i + v_3 i^2), i \in \{1, \ldots, n\}$.

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$ and define $\widetilde{\boldsymbol{a}} := \mathrm{alm}_n(\boldsymbol{a})$, $\widetilde{\boldsymbol{b}} := \mathrm{alm}_n(\boldsymbol{b}) \in \mathbb{R}^n$. The parameter matrices $A, B \in \mathbb{S}_n$ in (2.8) can be written as $A := \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top$, $B := \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top$, and hence the Almon DCC model specification is given by

$$Q_t = (\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top - \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot S + (\widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top) \odot (\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top) + (\widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot Q_{t-1} \tag{2.24}$$

and the associated constraints by

(IC) Identification constraints:
$$\widetilde{a}_1 > 0, \ \widetilde{b}_1 > 0, \quad \text{i.e.}$$
$$a_1 + \exp(a_2 + a_3) > 0, \ \ b_1 + \exp(b_2 + b_3) > 0. \tag{2.25}$$

(PC) Positivity constraints:
$$(\mathbf{i}_n\mathbf{i}_n^\top - \widetilde{\boldsymbol{a}}\widetilde{\boldsymbol{a}}^\top - \widetilde{\boldsymbol{b}}\widetilde{\boldsymbol{b}}^\top) \odot S \succ 0, \quad Q_0 \succeq 0. \tag{2.26}$$

In the non-targeted version of the Almon DCC model, the parameter matrix $C \in \mathbb{S}_n$ in (2.5) is parametrized by $\boldsymbol{c} \in \mathbb{R}^N$, $N = \frac{1}{2}n\,(n+1)$ via the assignment $C := \mathrm{math}(\boldsymbol{c})$, and hence the constraints on the parameters $\boldsymbol{a}$, $\boldsymbol{b} \in \mathbb{R}^3$, and $\boldsymbol{c} \in \mathbb{R}^N$ are:

(IC) Identification constraints:
$$\widetilde{a}_1 > 0, \ \widetilde{b}_1 > 0. \tag{2.27}$$

(PC) Positivity constraints:
$$\mathrm{math}(\boldsymbol{c}) \succ 0, \quad Q_0 \succeq 0. \tag{2.28}$$

(SC) Stationarity constraints:
$$\left|\widetilde{a}_i\widetilde{a}_j + \widetilde{b}_i\widetilde{b}_j\right| < 1, \ \ i, j \in \{1, \ldots, n\}, \ \ i \geq j. \tag{2.29}$$

2.5 The Almon shuffle DCC model

The Almon shuffle DCC specification is a variant of the Almon DCC model in which the different components of the process $\{\mathbf{r}_t\}$ are ordered so that the performance of the Almon parameterization is enhanced. Indeed, experience shows that the modeling performance of the Almon DCC prescription is much influenced by the ability of the Almon function to fit the entry values of the parameter vectors $\widetilde{\boldsymbol{a}}$ and $\widetilde{\boldsymbol{b}}$ that one would obtain by using an unrestricted rank deficient model with $r = 1$. This fit can be improved by first carrying out a reordering of the process components so that the vector entries are as monotonous as possible, hence fostering a good match between the typical profiles of Almon curves and the entry values of $\widetilde{\boldsymbol{a}}$ and $\widetilde{\boldsymbol{b}}$. The proposed reordering (shuffle) consists in arranging the components in descending order according to the magnitude of their projection onto the first principal component (the one that corresponds to the largest eigenvalue) computed using the unconditional covariance matrix of the available sample. Once the reordering is implemented, the Almon DCC model of the previous subsection is used. The results in Section 4 show that the Almon shuffle DCC model often exhibits a better performance than the Almon model.

2.6 The scalar DCC model

The scalar DCC model is by far the most widely used in the literature. In this case, the parameter matrices $A, B \in \mathbb{S}_n$ in (2.8) are of the form $A = a\mathbf{i}_n\mathbf{i}_n^\top$, $B = b\mathbf{i}_n\mathbf{i}_n^\top$, with $a, b \in \mathbb{R}$. The scalar DCC model specification is

$$Q_t = (1 - a - b)\,S + a\ \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top + b\ Q_{t-1}, \tag{2.30}$$

and the associated constraints have the form

(PC) Positivity constraints:
$$a \geq 0, \quad b \geq 0, \quad Q_0 \succeq 0. \tag{2.31}$$

(SC) Stationarity constraints:
$$a + b < 1. \tag{2.32}$$

As for all previously discussed DCC subfamilies, in the non-targeted version of the scalar model, the extra parameter matrix $C \in \mathbb{S}_n$ is intrinsically parametrized by $\boldsymbol{c} \in \mathbb{R}^N$, $N = \frac{1}{2}n\,(n+1)$ by defining $C := \mathrm{math}(\boldsymbol{c})$, and hence the associated constraints are (2.31) and (2.32) with the additional condition $\mathrm{math}(\boldsymbol{c}) \succ 0$.

We conclude this section with Table 2.1 that reports the number of parameters of the matrices A and B that are necessary in the different models for process dimensions going from five to thirty.

	$\boldsymbol{n}$	**5**	**10**	**15**	**20**	**25**	**30**
Hadamard	$n(n+1)$	30	110	240	420	650	930
Rank Deficient $(r=2)$	$2(2n-1)$	18	38	58	78	98	118
Rank Deficient $(r=1)$	$2n$	10	20	30	40	50	60
Almon	6	6	6	6	6	6	6
Scalar	2	2	2	2	2	2	2

Table 2.1: Number of parameters needed for the different DCC model parameterizations as a function of the process dimension n. These numbers represent exclusively the parameters obtained in the second stage estimation.

3 Constrained estimation of DCC model parameters

The purpose of this section is to present a constrained optimization scheme adapted to the quasi-maximum likelihood (QML) estimation of the DCC models previously described and to provide the main stages of its implementation; details are given in Section D of the TA. The QML approach stems from the assumptions that the innovations $\boldsymbol{\xi}_t$ appearing in (2.1) are normally distributed and that the conditional variance and correlation of the data generating process behave according to the Hadamard DCC equation.

3.1 The log-likelihood function and its gradient

In this subsection the log-likelihood function associated to the DCC model (2.1) with correlation dynamics determined by (2.8) and the estimation steps are briefly stated. The issues posed by the need to handle the parameter constraints are addressed in Subsection 3.2.

Let $\mathbf{r} = \{\mathbf{r}_1, \ldots, \mathbf{r}_T\}$ be a sample of size T of n-dimensional observations of the process $\{\mathbf{r}_t\}$ and let $\boldsymbol{\Theta} := (A, B) \in \mathbb{S}_n \times \mathbb{S}_n$ denote the parameters to be estimated, keeping in mind that this is performed after substituting the targeting estimator defined in (2.9) for the parameter S (in the sequel the symbol S stands for the targeting estimator instead of the unknown value). The log-likelihood function associated to the process (2.1) is

$$\log L\,(\boldsymbol{\Theta}; \mathbf{r}) = \sum_{t=1}^{T} l_t\,(\boldsymbol{\Theta}; \mathbf{r}_t)\,, \tag{3.1}$$

where

$$l_t\,(\boldsymbol{\Theta}; \mathbf{r}_t) = -\frac{1}{2}\left(n\log(2\pi) + \log\det\,(H_t) + \mathbf{r}_t^\top H_t^{-1}\mathbf{r}_t\right). \tag{3.2}$$

The dependence of l_t on $\boldsymbol{\Theta}$ is materialized through H_t, that explicitly relies on the set of parameters $\boldsymbol{\Theta}$ through Q_t since H_t is equal to $D_t Q_t^{*-1/2} Q_t Q_t^{*-1/2} D_t$. Recall that D_t depends on the parameters of the models used for the conditional volatilities and that are specified in the first stage of the DCC model construction. These parameters can be estimated consistently in a first stage as explained in [Eng02] and

in the sequel the symbol D_t in (3.2) and all the expressions derived from it stand for the corresponding estimated version. After this substitution, the expression (3.1) is strictly equivalent to the second stage log-likelihood function defined in [Eng02].

Consequently, the DCC QML (second stage) estimation problem consists of finding the parameter value $\widehat{\boldsymbol{\Theta}}$ that maximizes the log-likelihood (3.1) associated to a particular DCC model subjected to the constraints **(IC)**, **(SC)**, and **(PC)** associated to it. This can be carried out by using an iterative optimization method proposed in [CO14], adapted to the different versions of the DCC model defined in the previous section. For each of these versions, the matrices $A, B \in \mathbb{S}_n$ in (2.8) are functions of a parameter vector $\boldsymbol{\theta}$ that belongs to an intrinsic parameter space $\mathcal{P} \times \mathcal{P}$ that is specific to each model, and for which $\boldsymbol{\Theta} = \boldsymbol{\Theta}(\boldsymbol{\theta})$.

The computation of the estimator $\widehat{\boldsymbol{\Theta}}$ of $\boldsymbol{\Theta}$ via the constrained maximum likelihood optimization method that is used requires the gradient of the log-likelihood function of each of the DCC model specifications defined in Section 2. The analytical expressions of those gradients are given in Proposition D.1 in the technical appendix and in Subsection D.2 details are provided about the algorithmic implementation of Proposition D.1 via matrix recursions.

3.2 Constrained optimization using Bregman divergences

The constrained optimization method proposed to compute the QML estimator of the DCC model parameters is presented in this subsection. Handling the linear and nonlinear parameter constraints associated to each of the studied specifications complicates the optimization problem. Standard techniques, like Lagrange duality, often require to solve additional secondary optimization problems. This difficulty is circumvented by implementing a penalized optimization scheme that uses the so-called Bregman matrix divergences ([Bre67]) and that provides a solution using the primal space as working setup. This technique has been introduced in the context of machine learning (see for instance [DT07] and [KSD09b]) and has shown good performances in the estimation of the heavily parametrized VEC-GARCH model in [CO14].

3.2.1 Bregman divergences and constrained optimization problems

The Bregman matrix nearness measure is defined as follows: Let $X, Y \in \mathbb{S}_n$ and let $\phi : \mathbb{S}_n \longrightarrow \mathbb{R}$ be a strictly convex and differentiable function. The Bregman matrix nearness measure associated to ϕ is defined by $D_\phi(X, Y) := \phi(X) - \phi(Y) - \operatorname{tr}\left((\nabla\phi(Y))^\top (X - Y)\right)$. Different choices of ϕ lead to different measures, such as the squared Frobenius distance, the von Neumann divergence, and the Burg matrix divergence. The latter, also referred to as LogDet divergence or Stein's loss (mainly in the statistics literature, see [JS61]) is obtained out of the Burg entropy of the eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ of a positive definite matrix X, that is $\phi(X) := -\sum_{i=1}^n \log \lambda_i = -\log\det(X)$. The associated Bregman divergence defined in the space of positive definite matrices $X, Y \in \mathbb{S}_n^+$ is given by

$$D_M(X, Y) := \operatorname{tr}(XY^{-1}) - \log\det(XY^{-1}) - n. \tag{3.3}$$

In what follows, the denomination Bregman matrix divergence is used to refer to the Burg divergence.

Bregman divergences are particularly useful in optimization when dealing with positive (semi) definiteness constraints (abbreviated as positivity constraints **(PC)** in this paper). Consider the problem of minimizing a function $f(A)$ defined on a space of square matrices and subjected to the constraint $A \succ 0$. This is handled by iteratively solving optimization problems associated to penalized local models of the form

$$f_A^{(k)}(A) := f(A^{(k)}) + \nabla f(A^{(k)})(A - A^{(k)}) + \frac{1}{2}(A - A^{(k)})^T H(A^{(k)})(A - A^{(k)}) + D_M(A, A^{(k)}), \tag{3.4}$$

where $k \in \mathbb{N}$ is the label for the iteration, $\nabla f(A^{(k)})$ and $H(A^{(k)})$ are the gradient and the Hessian of the function f computed at the point $A^{(k)}$, respectively. The penalization term $D_M(A, A^{(k)})$ in (3.4) diverges when A approaches the set where the constraints are violated and forces the solution of the penalized local model to automatically lie in the constrained set. Proceeding this way the constrained optimization problem is reduced to a sequence of local unconstrained ones.

3.2.2 Bregman divergences for DCC model positivity constraints

The generic expressions of the Bregman divergences associated to the constraints imposed on the different DCC models under consideration are written down below in terms of the variables $\boldsymbol{\theta}$ of their intrinsic parameter space $\mathcal{P} \times \mathcal{P}$. In view of the different functional characters of the constraints considered, they can be classified into three groups.

Positive semidefinite (definite) constraints (PSDC): This group of constraints can be generically written as $M(\boldsymbol{\theta}) \succeq 0$, where $M : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{S}_q$, $q \in \mathbb{N}$, is a smooth map. In this case, the Bregman matrix divergence $D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{R}$ is given by

$$D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \operatorname{tr}\left(M(\boldsymbol{\theta}) \cdot M(\boldsymbol{\theta}^{(k)})^{-1}\right) - \log\det\left(M(\boldsymbol{\theta}) \cdot M(\boldsymbol{\theta}^{(k)})^{-1}\right) - q. \tag{3.5}$$

Nonlinear (NLPC) and linear (LPC) positivity constraints: The NLPC are specified by relations of the form $N(\boldsymbol{\theta}) > \mathbf{0}_q$, where $N : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{R}^q$ is a differentiable map and q is the number of components of the constraint. The corresponding entries of the divergence $D_N(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{R}^q$ are determined by the relation

$$D_N^i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = (D_N(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})_i = \frac{(N(\boldsymbol{\theta}))_i}{(N(\boldsymbol{\theta}^{(k)}))_i} - \log\frac{(N(\boldsymbol{\theta}))_i}{(N(\boldsymbol{\theta}^{(k)}))_i} - 1, \quad i = \{1, \ldots, q\}. \tag{3.6}$$

Linear positivity constraints (LPC) are a particular case of the previous ones, when N is a linear map. More specifically, if $\boldsymbol{\theta} \in \mathcal{P} \times \mathcal{P}$, we consider linear constraints of the form $L(\boldsymbol{\theta}) := \mathbf{f} - \sum_{i=1}^{2} C_{\theta_i}\theta_i > \mathbf{0}_m$, with $C_{\theta_i} \in \mathbb{M}_{m,P}$, $\theta_i \in \mathcal{P}$, $\mathbf{f} \in \mathbb{R}^m$, and $m \leq P \in \mathbb{N}$. The corresponding divergences are denoted by $D_L(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{R}^m$.

3.2.3 The local model for the DCC model, its gradient and associated Jacobian

As already explained in Section 3.2.1, the optimization algorithm is based on a sequence of penalized local functions that incorporate the Bregman divergences in order to ensure that the constraints are satisfied at each iteration. More specifically, the solution $\boldsymbol{\theta}^{(k+1)}$ of the local optimization problem after k iterations is defined by

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta} \in \mathcal{P} \times \mathcal{P}}{\arg\min}\ \tilde{f}^{(k)}(\boldsymbol{\theta}), \tag{3.7}$$

where the local objective function $\tilde{f}^{(k)}$ at $\boldsymbol{\theta}^{(k)}$ is given by

$$\begin{aligned}\tilde{f}^{(k)}(\boldsymbol{\theta}) :=& f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) \\ &+ \sum_{j=1}^{s_1} L_1^j D_{M_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{s_2} L_2^j \mathbf{i}_{q_j}^\top D_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{s_3} L_3^j \mathbf{i}_{m_j}^\top D_{L_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}).\end{aligned} \tag{3.8}$$

In the above expression, $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}^{(k)}; \mathbf{r})$ is the gradient of minus the log-likelihood function which is determined, in the DCC case, by the relation (D.1) in Proposition D.1. The symbol

$H^{(k)}$ denotes the Hessian of the function f computed at the point $\boldsymbol{\theta}^{(k)}$. The integers s_1, s_2, s_3 are the numbers of positive semidefiniteness, nonlinear, and linear constraints, respectively; the symbols $D_{M_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{R}$, $j \in \{1, \ldots, s_1\}$, $D_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{R}^{q_j}$, $j \in \{1, \ldots, s_2\}$, and $D_{L_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{R}^{m_j}$, $j \in \{1, \ldots, s_3\}$ denote the Bregman divergences defined in (3.5), (3.6), and the linear version of (3.6), respectively; $\mathbf{L}_1 \in \mathbb{R}^{s_1}, \mathbf{L}_2 \in \mathbb{R}^{s_2}, \mathbf{L}_3 \in \mathbb{R}^{s_3}$ are vectors whose components control the strength of the Bregman penalizations, and $\mathbf{i}_{q_j} \in \mathbb{R}^{q_j}$, $j \in \{1, \ldots, s_2\}$, $\mathbf{i}_{m_j} \in \mathbb{R}^{m_j}$, $j \in \{1, \ldots, s_3\}$ are vectors of ones.

The local optimization problem in (3.7) is solved by finding the value $\boldsymbol{\theta}_0$ for which $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta}_0) = 0$. The complete expressions of the gradient and associated Jacobian $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ of the local model (3.8) for the different DCC specifications are provided in Section E of the technical appendix.

4 Empirical study

The goal of this section is to report on the results of experiments carried out in order to compare the empirical performances of the different DCC models described in Section 2. For that purpose a dataset is selected and those models are estimated in various dimensions (from 5 to 30) using the algorithm presented in Section 3. Their in-sample fit is evaluated with the Akaike information criterion (AIC) in Subsection 4.2, which also contains a discussion about the heterogeneity and other features of the point estimates. Statistical tests related to the out-of-sample performances of the different models are described in Subsection 4.3 and their results are discussed in Subsection 4.4.

4.1 Dataset and competing models

The dataset used for the empirical study consists of the daily closing price quotes of the thirty components included in the Dow Jones Industrial Average Index (DJIA) as of October 2013. The data are downloaded from the Yahoo Finance database. The Yahoo tickers of the stocks used in the study are AA, AXP, BA, BAC, CAT, CSCO, CVX, DD, DIS, GE, HD, HPQ, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, PFE, PG, T, TRV, UNH, UTX, VZ, WMT, XOM. The dataset has been prepared adjusting the quotes with respect to stock splits and dividend payments and the dates at which at least one of the constituents was not quoted were removed. The price quotes are taken from August 25th, 1998 to August 1st, 2013. The resulting sample contains 3750 observations. The first 3000 quotes (August 25th, 1998 - August 9th, 2010) are used for model estimation and the last 750 are kept for out-of-sample testing. Another similar dataset covering a different period has been processed and studied; the corresponding results are available in Section G of the technical appendix.

Data preprocessing using the Capital Asset Pricing Model (CAPM) In order to account for the common dynamical factor that influences the assets under consideration, a static unconditional CAPM one-factor model is used for each asset $i \in \{1, 2, \ldots, n\}$. It is of the form $Y_{i,t} = \alpha_i^\circ + \beta_i^\circ X_t + Z_{i,t}$, where for each time index t, $Y_{i,t}$ is the log-return of asset i, X_t is the value of the chosen common factor, $Z_{i,t}$ is the regression error term, and $\alpha_i^\circ, \beta_i^\circ$ are the intercept and slope coefficients, respectively.

In the DCC empirical experiments presented in this section, the CAPM regression is used in the following way: let T_{est} and T_{out} be the sample lengths taken for in-sample estimation and out-of-sample testing, respectively, and let $T := T_{est} + T_{out}$ be the total number time series observations. The log-returns of the S&P500 index are used as the common factor X_t. The CAPM regression is estimated by ordinary least-squares for each asset i using the observations $t \in \{1, \ldots, T_{est}\}$, and the OLS residuals $Z_{i,t}$ are stored. Then, following the same approach as in Chapter 8 of [Eng09], these residuals $\mathbf{Z}_t := (Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t})^\top$ are used as the observed returns $\mathbf{r}_t$ appearing in (2.1). Next, the two-stage estimation of the DCC models is applied, as explained in Section 2; more specifically, in the first stage, GARCH(1,1) models are fit to the components of $\mathbf{Z}_t$, and in the second one the different DCC

models are estimated. The complete model (with the CAPM step) is thus what is called "FACTOR DCC" in Chapter 8 of [Eng09], except that the factor (that is, the S&P 500 index) volatility and its correlations with the thirty assets are not modeled. Instead of simple GARCH(1,1) models as in (2.2), other prescriptions can be used that incorporate, for example, an asymmetry effect.

In order to perform the out-of-sample analysis, the corresponding out-of-sample residual returns $\mathbf{Z} := \{\mathbf{Z}_{T_{est}+1}, \ldots, \mathbf{Z}_T\}$ are computed using the relations

$$Z_{i,t} = Y_{i,t} - \widehat{\alpha_i^\circ} - \widehat{\beta_i^\circ} X_t, \quad i \in \{1, \ldots, n\}, \quad t \in \{T_{est}+1, \ldots, T\}, \tag{4.1}$$

where $\widehat{\alpha_i^\circ}$ and $\widehat{\beta_i^\circ}$ are obtained using the T_{est} in-sample observations and subsequently kept for the out-of-sample assessment of the empirical performances of the DCC models. When generating the out-of-sample forecasts of the H_t matrices of each DCC model, the values of the parameter estimates obtained using only the in-sample observations are also kept.

The CAPM based data preprocessing step is not absolutely necessary for the models to show good performance; its relevance depends on the nature of the data and it is up to the practitioner to carry out this step. The empirical study reported in this section has also been implemented without CAPM preprocessing and the conclusions drawn from these results are broadly similar to those drawn from the results presented in the sequel.

The competing models Results are reported for six different DCC model parameterizations, namely: **(i)** the Hadamard DCC, **(ii)** the rank deficient DCC with rank $r = 2$, **(iii)** the rank deficient DCC with rank $r = 1$, **(iv)** the Almon DCC, **(v)** the Almon shuffle DCC, and **(vi)** the scalar DCC. Some of these models are particular cases of others according to the inclusion relations represented in Figure 1.

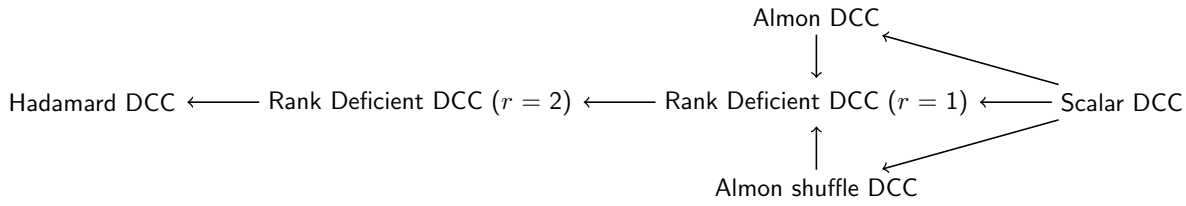

Figure 1: Inclusion hierarchy of the models. The symbol $B \longleftarrow A$ means "A is a particular case of B".

4.2 In-sample results

The model estimations are performed in various dimensions ranging from 5 up to 30 in such a way that for the n-dimensional case the first n assets are picked in the DJIA dataset arranged in alphabetical order. The estimations are carried out conditionally on the estimated standardized ("degarched") residuals and the targeting estimator of S, which are the same for all the models. Hence the computed optimal values of the log-likelihood functions are fully comparable across models.

Goodness-of-fit comparisons A consequence of the inclusion relations shown in Figure 1 is that the optimal log-likelihood values obtained in the estimation of these models for a given sample need to be ordered accordingly. More specifically, the optimal value resulting from the estimation procedure for the Hadamard DCC model has to be the smallest among them all, since it is minus the log-likelihood function which is minimized, and the other models need to respect the hierarchy established by the diagram in Figure 1. In Table 4.2 these values and the associated AIC statistics are reported, with

AIC defined as AIC := $(-2\log L + k\log T_{est})/T_{est}$, where T_{est} is the size of the sample reserved for estimation, $\log L$ is the obtained maximal value of the log-likelihood function associated to the model and the sample used for estimation, and k is the number of parameters of the model. A lower AIC value indicates a better trade-off between the quality of the model fit and the number of parameters used to achieve it.

The values reported in the table reveal that the Hadamard DCC model has, as should be the case, the smallest minus log-likelihood values after convergence of the algorithms, and that the hierarchy of models in terms of the maximized values is respected. At the same time, the rank deficient DCC models with rank one (in four cases) and rank two (in two cases) exhibit the best trade-off between fit and parsimony according to the AIC statistics. In all cases, the Hadamard models have the largest AIC values, which is not a big surprise given that they have many more parameters than the other models (see Table 2.1). It is interesting to notice that the differences in AIC values between the Hadamard model and all the other models are clearly larger than the differences between the other models. If the six models are ranked by attributing the score 6 to the worst AIC fitting model (the Hadamard), the score 5 to the next, up to 1 to the best performing model, and these scores are added up through the six dimensions, the best performing model is the rank one deficient model, the second best model is the rank two deficient model, the third is the Almon shuffle, and the fourth the scalar. These results give a first indication that the rank one deficient and Almon shuffle models seem to be worth using in practice, in addition to the scalar model.

n		Scalar	Almon	Almon Shuffle	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Hadamard
5	-log$L$	-13.4686	-13.4690	-13.4713	-13.4728	-13.4735	**-13.4736**
	AIC$^{\text{rank}}$	-26.9258^{3}	-26.9239^{5}	-26.9286^{2}	**-26.9289**^{1}	-26.9250^{4}	-26.9173^{6}
10	-log$L$	-27.8050	-27.8059	-27.8084	-27.8149	-27.8178	**-27.8186**
	AIC$^{\text{rank}}$	-55.5888^{4}	-55.5878^{5}	-55.5928^{2}	**-55.5964**^{1}	-55.5902^{3}	-55.5439^{6}
15	-log$L$	-41.8992	-41.9006	-41.9026	-41.9164	-41.9235	**-41.9266**
	AIC$^{\text{rank}}$	-83.7671^{5}	-83.7671^{4}	-83.7712^{3}	**-83.7829**^{1}	-83.7784^{2}	-83.6633^{6}
20	-log$L$	-56.4619	-56.4629	-56.4662	-56.4860	-56.4969	**-56.5068**
	AIC$^{\text{rank}}$	-112.8824^{4}	-112.8819^{5}	-112.8885^{3}	**-112.9053**^{1}	-112.9018^{2}	-112.6935^{6}
25	-log$L$	-71.0132	-71.0148	-71.0172	-71.0357	-71.0523	**-71.0724**
	AIC$^{\text{rank}}$	-141.9750^{5}	-141.9757^{4}	-141.9805^{3}	-141.9880^{2}	**-141.9893**^{1}	-141.6614^{6}
30	-log$L$	-85.9028	-85.9063	-85.9065	-85.9347	-85.9565	**-85.9935**
	AIC$^{\text{rank}}$	-171.7443^{5}	-171.7487^{4}	-171.7490^{3}	-171.7694^{2}	**-171.7743**^{1}	-171.3070^{6}
	AIC score$^{\text{rank}}$	26^{4}	27^{5}	16^{3}	**8**^{1}	13^{2}	36^{6}

Table 4.2: Normalized values of minus the log-likelihood function (-log$(L)/T_{est}$), and associated AIC statistics. The smallest values of minus the log-likelihood function are displayed in black bold. Exponents on the AIC row indicate the rank of the model from 6 (the worse) to 1 (the best). The "AIC score" row at the bottom contains aggregated ranks by models. Figures in red point to the model that exhibits the lowest AIC value.

Features of the point estimates Since non-scalar DCC models allow for heterogeneity in the parameters that drive the conditional correlations, some information on their estimates is of interest. Table 4.3 reports on the mean, standard deviation, minimum, and maximum of the estimates of the elements of the matrices A and B of the DCC process written as in (2.8), for each model and dimension. If these matrices are functions of intrinsic parameters of lower dimension than in the Hadamard case, then the matrices implied by the intrinsic parameter estimates are computed and the corresponding statistics are obtained. Obviously, for the scalar model, the mean, maximum, and minimum are equal to the point estimate of the corresponding scalar parameters. The following general conclusions can be drawn:

Matrix A

n		MacGyver	Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	mean (median)	0.0103 (0.0086)	0.0063	0.0093	0.0071	0.0092	0.0064	0.0065
	std	0.0090	-	0.0102	0.0054	0.0106	0.0039	0.0042
	min	0.0000	0.0063	0.0004	0.0021	-0.0000	0.0026	0.0026
	max	0.0432	0.0063	0.0493	0.0276	0.0520	0.0194	0.0201
10	mean (median)	0.0118 (0.0101)	0.0060	0.0066	0.0065	0.0063	0.0060	0.0062
	std	0.0089	-	0.0036	0.0025	0.0028	0.0003	0.0013
	min	0.0000	0.0060	-0.0015	0.0025	0.0012	0.0056	0.0045
	max	0.0432	0.0060	0.0189	0.0124	0.0122	0.0071	0.0112
15	mean (median)	0.0101 (0.0079)	0.0044	0.0047	0.0048	0.0045	0.0044	0.0045
	std	0.0082	-	0.0038	0.0029	0.0035	0.0008	0.0014
	min	0.0000	0.0044	-0.0032	0.0005	-0.0022	0.0021	0.0026
	max	0.0432	0.0044	0.0243	0.0193	0.0197	0.0055	0.0100
20	mean (median)	0.0099 (0.0074)	0.0034	0.0033	0.0037	0.0031	0.0035	0.0035
	std	0.0089	-	0.0026	0.0022	0.0023	0.0005	0.0005
	min	0.0000	0.0034	-0.0039	0.0004	-0.0030	0.0027	0.0021
	max	0.0535	0.0034	0.0162	0.0144	0.0110	0.0052	0.0042
25	mean (median)	0.0093 (0.0069)	0.0031	0.0029	0.0032	0.0030	0.0031	0.0031
	std	0.0093	-	0.0023	0.0018	0.0017	0.0005	0.0003
	min	0.0000	0.0031	-0.0027	0.0008	-0.0010	0.0022	0.0021
	max	0.0545	0.0031	0.0155	0.0130	0.0098	0.0046	0.0035
30	mean (median)	0.0093 (0.0070)	0.0027	0.0025	0.0027	0.0025	0.0026	0.0027
	std	0.0092	-	0.0023	0.0019	0.0021	0.0005	0.0002
	min	0.0000	0.0027	-0.0040	0.0004	-0.0021	0.0020	0.0020
	max	0.0592	0.0027	0.0138	0.0112	0.0131	0.0047	0.0031

Matrix B

n		MacGyver	Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	mean (median)	0.8748 (0.9673)	0.9815	0.9728	0.9694	0.9725	0.9819	0.9788
	std	0.2433	-	0.0121	0.0170	0.0126	0.0106	0.0057
	min	0.0006	0.9815	0.9461	0.9265	0.9436	0.9550	0.9672
	max	0.9945	0.9815	0.9928	0.9944	0.9937	0.9956	0.9881
10	mean	0.8748 (0.9647)	0.9801	0.9742	0.9667	0.9733	0.9785	0.9772
	std	0.2617	-	0.0088	0.0228	0.0148	0.0051	0.0022
	min	0.0000	0.9801	0.9478	0.8785	0.9197	0.9641	0.9743
	max	0.9949	0.9801	0.9928	0.9962	0.9962	0.9865	0.9842
15	mean (median)	0.8874 (0.9658)	0.9827	0.9797	0.9737	0.9788	0.9822	0.9815
	std	0.2309	-	0.0070	0.0158	0.0141	0.0009	0.0049
	min	0.0000	0.9827	0.9626	0.9258	0.9297	0.9808	0.9653
	max	0.9975	0.9827	0.9942	0.9984	0.9970	0.9851	0.9878
20	mean (median)	0.8769 (0.9672)	0.9866	0.9854	0.9806	0.9857	0.9862	0.9851
	std	0.2415	-	0.0038	0.0104	0.0073	0.0013	0.0029
	min	0.0000	0.9866	0.9758	0.9434	0.9601	0.9840	0.9805
	max	0.9975	0.9866	0.9951	0.9985	0.9985	0.9900	0.9942
25	mean (median)	0.8752 (0.9724)	0.9862	0.9857	0.9816	0.9848	0.9860	0.9853
	std	0.2490	-	0.0026	0.0094	0.0076	0.0007	0.0028
	min	0.0000	0.9862	0.9791	0.9465	0.9485	0.9851	0.9812
	max	0.9975	0.9862	0.9925	0.9964	0.9977	0.9881	0.9950
30	mean (median)	0.8776 (0.9718)	0.9866	0.9858	0.9836	0.9844	0.9867	0.9858
	std	0.2391	-	0.0020	0.0099	0.0089	0.0021	0.0029
	min	0.0000	0.9866	0.9781	0.9383	0.9420	0.9839	0.9815
	max	0.9975	0.9866	0.9920	0.9985	0.9974	0.9943	0.9964

Table 4.3: Estimated parameter matrices A and B of (2.8). Mean, median, standard deviation, minimum and maximum of the entries of the estimated matrices are reported. "MacGyver" stands for the method with this name introduced in [Eng08] based on the use of bivariate scalar DCC models; in this case the median values of the estimates are reported between parentheses.

- The A mean values decrease towards zero in all models as the dimension n increases: they are approximately divided by a factor of 2 to 3 in dimension 30 relative to dimension 10. This is a well-known phenomenon for the scalar DCC model, already mentioned in [Eng02] and discussed in [PESS14]. The latter paper shows by a simulation study that the a parameter is subjected to a downward bias and that variance targeting is responsible for that. Our results suggest that the same problem occurs in non-scalar models. Moreover, the mean values are very close to each other across the different models. This is less the case in dimension five, a result probably due to the arbitrary selection of the assets. Taking this case aside, this suggests that the scalar model estimates the average value of the elements of A, and that all models are likely to be subjected to the downward bias problem. Further insight on this issue could be obtained by a simulation study which is left for further research.

- The A standard deviations (which measure the degree of heterogeneity between the elements of A) decrease as the dimension increases (though not monotonically, see dimensions 10 and 15, depending on the model). The decrease is more pronounced between dimensions 5 and 15 than between dimensions 20 and 30.

- For every dimension, the largest degree of heterogeneity in A among non-scalar models occurs for the Hadamard case (with one minor exception for dimension 5 where it is the second largest). The rank deficient models have a slightly smaller degree of heterogeneity than Hadamard, while the Almon models exhibit much less heterogeneity than the previous three models as a consequence of their more parsimonious parameterization.

- The B mean values increase slightly towards unity until dimension 20, and stay approximately the same for dimensions 20, 25, and 30.

- The B standard deviations decrease globally (comparing dimensions 10 and 30), with some intermediate ups and downs, as the dimension increases.

Table 4.3 also provides the statistics of the estimates coming from the MacGyver method proposed in [Eng08]. For dimension 30, this consists in estimating all the bivariate scalar DCC models that can be obtained using distinct combinations (there are 435 of them) of the 30 assets in question, thus getting as many point estimates of a and b. For each dimension smaller than 30 (except 5), the estimates for the corresponding group of assets are obtained in this way (e.g. the distinct pairs of assets among the first ten assets are used for dimension 10, which provides 45 estimates that are a subset of the 435 obtained for dimension 30). For dimension 5 too little combinations are available and hence six groups of five assets are formed, (1-5, 6-10, until 26-30) which yield in total 60 estimates for the six groups. For each dimension, the MacGyver estimates exhibit much more heterogeneity than the different DCC models do. This is a consequence of the positive semidefiniteness constraints imposed on the DCC models to which the MacGyver estimator is not exposed. Additionally, unusual values arise sometimes in the MacGyver case in comparison with the DCC estimates; more specifically, there are values of b close to zero (see the minimum MacGyver values) and of a farther away from zero (see the MacGyver maximum values). The mean values of the MacGyver estimates of a are therefore larger than the corresponding medians (reported in the tables), and the reverse is true for the b estimates. The median values of the a and b estimates are hardly influenced by the dimension.

Rank one deficient and Almon shuffle point estimates Given that the AIC statistics favor these models, more information on the estimates of the corresponding parameters is of interest. Figure 2 plots the values of the entries of the parameter vector $\boldsymbol{a}$ for both models in dimension 30, ordered according to the Almon shuffle prescription. The figure also shows the corresponding estimate implied by the scalar model, which is the square root of the parameter a in (2.30). The figure illustrates the flexibility

of the rank one deficient and Almon models with respect to the scalar one. The pattern of estimates of the entries of $\boldsymbol{a}$ of the Almon model is concave, whereas for $\boldsymbol{b}$ (the corresponding figure is provided in the technical appendix) it is convex.

For the rank one deficient model, the estimates are naturally more fluctuating. The scalar model estimates correspond roughly to the average values of the individual estimates of the other models.

The estimates of the entries of $\boldsymbol{a}$ and $\boldsymbol{b}$ in the Almon shuffle model are computed by using the Almon function defined in Section 2.4 using the estimates of the three parameters of each Almon function for the dimensions considered (the relevant tables containing them are provided in the technical appendix). As the dimension increases, for the function defining $\boldsymbol{a}$, the estimates of the parameters in the exponential function (a_2 and a_3) tend to zero, and the estimates of the constant term (a_1) tend to minus one. This implies that the entries of $\boldsymbol{a}$ tend to zero and are more homogeneous in higher dimensions. These trends seem to saturate beyond dimension 15. Regarding the parameters of the Almon function defining $\boldsymbol{b}$, the three parameters tend to zero as the dimensions increases, so that the elements of $\boldsymbol{b}$ tend to one from below.

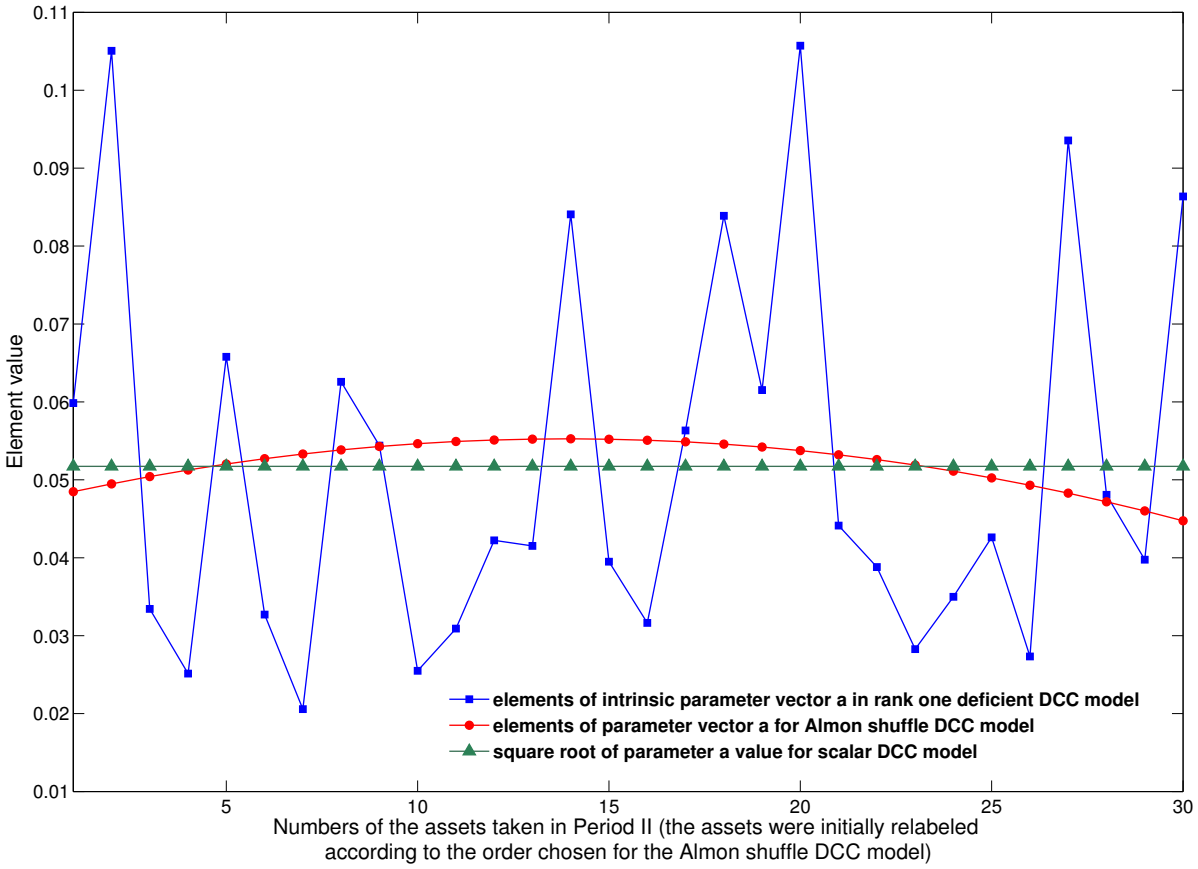

Figure 2: Estimates of the entries of the parameter $\boldsymbol{a}$ of the rank one deficient and Almon shuffle models for the thirty assets case in the order determined by the Almon shuffle estimation.

Computational effort and starting values Table 4.4 provides the time and the number of gradient calls necessary to carry out the estimation of the various DCC models in different dimensions. The figures provided for the non-scalar models are relative to the computational effort necessary to estimate the scalar model in the corresponding dimension. An important consideration that should be kept in

Relative computation time/Relative number of gradient calls.

n		Almon	Almon Shuffle	Rank Deficient ($r = 1$)	Rank Deficient ($r = 2$)	Hadamard
5	time	0.6599	0.7783	0.7828	6.5984	1.0486
	grad calls	0.6613	0.7742	0.6452	3.1935	0.6774
10	time	2.5637	5.4408	1.7209	28.3579	2.8449
	grad calls	0.7500	1.2333	0.9333	3.2667	0.9000
15	time	3.0451	16.9112	6.7186	14.4835	2.6498
	grad calls	1.1190	4.8095	1.7619	2.5952	1.0952
20	time	1.1361	0.9467	2.7956	2.6302	0.6662
	grad calls	1.2449	1.0816	1.8163	1.8571	0.7959
25	time	1.2971	1.8973	2.0043	2.6563	0.5554
	grad calls	1.4286	1.7959	1.7143	2.1837	0.8163
30	time	1.8630	1.9192	1.9769	1.9992	0.3627
	grad calls	1.9792	2.3750	2.0625	1.8958	0.7083

Table 4.4: Time and number of gradient calls necessary to carry out the estimation of the various DCC models under consideration in different dimensions. The figures are relative to the computational effort necessary to estimate the scalar model in the corresponding dimension. Several considerations sometimes make those figures not directly comparable: see explanations in page 17.

mind and that makes those figures not fully comparable is that due to the different sensitivities of the log-likelihood functions of the different models to variations in the intrinsic parameters, the stopping tolerances in the optimization algorithm had to be tuned for each different parameterization in order to secure reasonable results (10^{-4} for the Hadamard case and rank deficient parameterizations; for the Almon models, 10^{-4} in dimensions up to fifteen and 10^{-5} for $n = 20, 25, 30$). Consequently, the relative times reported in the table should not be interpreted as comparable numerical efficiency measures, but rather as an indication that the estimation of the more complex models is feasible in terms of computation time when compared to the scalar model.

Additionally, the figures in Table 4.4 are influenced by the choice of initial values in the implementation of the optimization algorithm. Our experience shows that a good approach to this question consists of using the estimated values for the scalar model as initial values for the estimation of the non-scalar ones. More precisely, the scalar model is estimated (for each dimension) using the same initial condition ($a = 0.2$, $b = 0.7$) and then, the resulting A and B matrices are fed as initial conditions to the optimization algorithm for the estimation of each of the other models. This way to proceed is reasonable in view of the model hierarchy depicted in Figure 1 and has important consequences at the time of avoiding the difficulties caused by the non convex character of the optimization problem that needs to be solved.

It is interesting to observe that the built-in nonlinearities of some of the non-scalar parameterizations (exponential for Almon and quadratic for rank deficient models) have sometimes a bigger impact on the computational effort than the dimension of the intrinsic parameter space (compare for example the values exhibited by the Hadamard parameterization with those associated to rank two deficient models).

The main conclusion that can be drawn from the data in Table 4.4 is that from a computational effort standpoint, non-scalar models are quite affordable with respect to the scalar model in the dimensions

considered.

4.3 Out-of-sample specification tests

In this section are presented the specification tests used in order to assess the out-of-sample one-step ahead forecasting performance of the competing DCC models. The discussion of the results follows in the next subsection. The first test is based on the use of multivariate variance standardized returns and the next three on the use of portfolio returns.

4.3.1 Model confidence set based on correlation loss functions

The different models are compared by computing the model confidence set (MCS) of [HLN03, HLN11] with the following loss function constructed using the GARCH standardized returns in (4.3) and the conditional correlation matrices implied by the parameter estimates of the models:

$$d_t := \frac{2}{n(n-1)} \sum_{i<j=2,\dots,n} \left(\varepsilon_{i,t}\varepsilon_{j,t} - \rho_{ij,t}\right)^2, \tag{4.2}$$

where $\rho_{ij,t}$ is the (i,j)-entry of the model dependent conditional correlation matrix R_t introduced in (2.4). The loss function is based on the fact that if (2.1) is a correct specification, the GARCH standardized returns

$$\boldsymbol{\varepsilon}_t := D_t^{-1/2}\mathbf{r}_t \tag{4.3}$$

have the correlation matrix R_t defined in (2.4). The results are provided in Table 4.5 which shows, for a given number of assets, the order of exclusion of models from the confidence set: 6 means that the corresponding model was the first to be excluded (with its p-value underneath), 1 means that the corresponding model was the last one. The MCS at 90% confidence level is identified by the set of bold red figures and at 95% by the union of the figures in bold red and bold black.

4.3.2 Tests based on portfolio returns

The performances of the competing DCC models can be compared indirectly by running tests on portfolios constructed using the assets whose returns are modeled. Let $\mathbf{w}_t \in \mathbb{R}^n$ denote a vector of portfolio weights at date t, let $p_t = \mathbf{w}_t^\top \mathbf{r}_t$ be the portfolio return, and $\sigma_{p,t}^2 = \mathbf{w}_t^\top H_t \mathbf{w}_t$ the corresponding variance, where H_t is the relevant conditional covariance matrix of $\mathbf{r}_t$ (see (2.1)). Two kinds of portfolios are constructed:

- The minimum variance portfolio (MVP), defined by choosing a weight vector $\mathbf{w}_t$ that minimizes $\mathbf{w}^\top H_t \mathbf{w}$ subjected to the constraint $\mathbf{i}_n^\top \mathbf{w}_t = 1$. The solution of this problem is given by $\mathbf{w}_t = H_t^{-1}\mathbf{i}_n / (\mathbf{i}_n^\top H_t^{-1} \mathbf{i}_n)$. This expression is used to construct the sequence of variance minimizing portfolios associated to each model.
- The equally weighted portfolio (EWP), implied by the weight vector $\mathbf{w}_t := \mathbf{i}_n/n$ for each date t.

Three tests are considered, based on the observation that under a correct specification of the type (2.1), the standardized portfolio return $y_t = \mathbf{w}_t^\top \mathbf{r}_t / \sqrt{\mathbf{w}_t^\top H_t \mathbf{w}_t}$, has unconditional variance equal to one. The tests assess the validity of different hypotheses for the series $\widehat{y}_t = \mathbf{w}_t^\top \mathbf{r}_t / \sqrt{\mathbf{w}_t^\top \widehat{H}_t \mathbf{w}_t}$ constructed using the one-step ahead forecast of the conditional covariance matrices $\widehat{H}_t$ implied by each of the estimated models under consideration.

Engle-Colacito regression test ([EC06]): It is constructed by estimating the regression

$$\widehat{y}_t - 1 = \lambda + u_t, \qquad t \in \{T_{est}+1, \dots, T\}, \tag{4.4}$$

where u_t is the error term and λ is the intercept coefficient. The test assesses the null hypothesis that λ is equal to 0 using a heteroskedasticity and autocorrealtion consistent (HAC) t-statistic ([And91]). The results are presented in Tables 4.6 and 4.7 that contain the HAC t-statistics and their p-values.

Model confidence set (MCS) based on the predictive ability for squared portfolio returns: The one-step ahead predictive ability of the models is evaluated by computing model confidence sets using the loss function

$$d_t := \left((\mathbf{w}_t^\top \mathbf{r}_t)^2 - \mathbf{w}_t^\top \widehat{H}_t \mathbf{w}_t \right)^2. \tag{4.5}$$

The results of this procedure are provided in Tables 4.8 and 4.9. The information that they contain is organized in the same way that was already described at the end of Section 4.3.1.

Value-at-Risk backtesting and the dynamic quantile test (HIT test): The backtesting of the dynamical structure of the Value-at-Risk (VaR) violations, that is, the occurrences for which $\mathbf{w}_t^\top \mathbf{r}_t < \mathrm{VaR}_\alpha(t)$, can be used as a tool to assess the performance of the model used to forecast the VaR. Let $\mathrm{HIT}_\alpha(t) := \mathbf{1}_{\{\mathbf{w}_t^\top \mathbf{r}_t < \mathrm{VaR}_\alpha(t)\}}$. Under a correct model specification, $E\left[\mathrm{HIT}_\alpha(t)\right] = \alpha$ and the random variables $\{\mathrm{HIT}_\alpha(t)\}_{t \in \{T_{est}+1,\ldots,T\}}$ are serially independent and independent of other elements in the conditioning information set, like the $\mathrm{VaR}_\alpha(t)$ itself. These features are tested by using the HIT-test defined in [EM04], which is a F-test on the regression

$$\mathrm{HIT}_\alpha(t) - \overline{\mathrm{HIT}}_\alpha = \lambda + \beta_1 \mathrm{HIT}_\alpha(t-1) + \cdots + \beta_l \mathrm{HIT}_\alpha(t-l) + \beta_{l+1} \mathrm{VaR}_\alpha(t) + u_t, \tag{4.6}$$

where $\overline{\mathrm{HIT}}_\alpha = \dfrac{1}{T_{out}} \sum_{t=T_{est}+1}^{T} \mathrm{HIT}_\alpha(t)$ and l is the chosen number of lags. The null hypothesis is that all the regression coefficients, including the intercept, are equal to zero. In order to compute $\mathrm{VaR}_\alpha(t)$, use is made of the fact that the forecasted portfolio returns $\widehat{p}_t$ are Gaussian with zero mean and estimated variance $\widehat{\sigma}_{p,t}^2 = \mathbf{w}_t^\top \widehat{H}_t \mathbf{w}_t$ and hence $\mathrm{VaR}_\alpha(t) = \Phi^{-1}(\alpha)\widehat{\sigma}_{p,t}$, where Φ is the standard normal cumulative distribution function.

The results of this test with lag order $l = 5$ for the two types of portfolios are described in Tables 4.10 and 4.11. For each portfolio cardinality, the tables contain the hit averages $\overline{\mathrm{HIT}}_\alpha$ for the confidence levels 1, 5, and 10%, as well as the corresponding p-values of the F-test on the regression (4.6).

4.4 Results of the out-of-sample specification tests

The four tests presented in the previous subsection are applied in several dimensions ranging from 5 up to 30 in such a way that for the n-dimensional case, the first n assets in the DJIA dataset are picked up arranged in alphabetical order. The conclusions about the results of the tests are presented below.

1) MCS for the correlation of the standardized returns (Table 4.5): The first observation that can be made is that the size of the MCS goes down when the dimension increases. The MCS contains 5 or 6 models for dimensions up to 15. It contains one model for $n = 20$, and two models for $n = 25$ and 30. These models are always Almon or Almon shuffle. The Almon shuffle model performs relatively well in all dimensions (with a score of 8), distantly followed by the Almon and rank one deficient models (scores 20), the scalar model having the fourth score (22). We emphasize that these MCS results do not imply that the different models produce on average very different correlation forecasts (when averaging is done with respect to all 435 correlation series for each model). An advantage of the MCS approach is that it discards a model if it is found to be significantly (at a pre-specified level) less well performing than the other models in the set.

2) Tests based on the use of portfolio returns.

- **Engle-Colacito regression test (Tables 4.6-4.7)**: The test results are the same in most cases across models for each dimension, that is, rejections (or non-rejections) at the level of 5% occur

MCS for the correlation of the standardized returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	5	**4**	**1**	**3**	6	**2**
	p-value	0.008	0.073	1.000	0.228	0.005	0.228
10	Position	**5**	**3**	**1**	**4**	**6**	**2**
	p-value	0.190	0.433	1.000	0.260	0.190	0.433
15	Position	**3**	**5**	**4**	**6**	**2**	**1**
	p-value	0.242	0.127	0.242	0.061	0.811	1.000
20	Position	3	6	5	4	2	**1**
	p-value	0.003	0.003	0.003	0.003	0.026	1.000
25	Position	3	6	5	4	**2**	**1**
	p-value	0.002	0.000	0.001	0.002	0.074	1.000
30	Position	3	6	4	5	**2**	**1**
	p-value	0.006	0.000	0.006	0.001	0.394	1.000
	Score	22	30	20	26	20	8

Table 4.5: Model confidence sets (MCS) constructed using the loss function (4.2) based on the correlation of the standardized returns defined in (4.3). For each model and dimension, the integer value in the first row indicates the order of elimination of the model from the MCS (6 stands for the first eliminated model, 5 for the second eliminated model, and so on). In the second row we report the p-value of the test leading to the decision of eliminating the given model from the MCS. The set of integer values printed in bold red identifies the MCS at the confidence level of 90%. The union of integer values printed in bold black and bold red identifies the MCS at the confidence level of 95%. The score of each model in the last row is the sum of the integer values of the six dimensions.

for all models, with the exception of dimension 25 for MVP (no rejection for scalar and Almon shuffle) and EWP (no rejection for the Hadamard model). However, when differences of this kind happen, the p-values are very close to 5%.

- **MCS for the equal predictive ability (EPA) of squared portfolio returns (Tables 4.8-4.9)**: The results do not favor systematically a particular model. For minimum variance portfolios, the model confidence sets at the 95% level (90% in a few cases) include all the models. The reported model scores favor the rank one deficient model. For equally weighted portfolios, the model confidence sets do not include all six models, and they only include the scalar model in dimensions 10, 15, and 20.

- **VaR backtesting and the dynamic quantile test (HIT test, Tables 4.10-4.11)**: For each dimension, the percentages of VaR violations (at 1, 5, and 10% confidence) and the p-values of the F-tests of independence are in most cases similar across models.

The results of the three tests based on portfolio returns clearly indicate that the different DCC models have more or less the same performance in terms of out-of-sample specification tests. No model dominates systematically the others, and no model is systematically dominated. This may be due to the fact that the different DCC models, and consequently these test results, rely on the same univariate GARCH models for the conditional variances. These conditional variances appear as diagonal elements in the matrices H_t that are used at the time of running the tests, and they affect the results also through the covariances deduced from the correlations and standard deviations. Consequently, the test results

Engle-Colacito regression test for the minimum variance portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	t-stats	-3.81*	-3.77*	-3.58*	-3.70*	-3.87*	**-3.49***
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
10	t-stats	0.39	0.30	0.38	**0.25**	0.33	0.29
	p-value	0.70	0.76	0.71	0.81	0.74	0.77
15	t-stats	**0.38**	0.66	0.40	0.49	0.43	0.41
	p-value	0.71	0.51	0.69	0.63	0.67	0.68
20	t-stats	0.58	1.01	0.72	0.82	0.62	**0.50**
	p-value	0.56	0.31	0.47	0.41	0.54	0.62
25	t-stats	1.96	2.77*	2.29*	2.31*	2.06*	**1.83**
	p-value	0.05	0.01	0.02	0.02	0.04	0.07
30	t-stats	2.58*	3.63*	2.73*	3.04*	2.73*	**2.35***
	p-value	0.01	0.00	0.01	0.00	0.01	0.02

Table 4.6: Results of the Engle-Colacito regression test. The t-stat values refer to the intercept λ of the Engle-Colacito regression (4.4) obtained using a HAC estimator. The p-value can be used to test the null hypothesis that $\lambda = 0$. The symbol * (respectively **) indicates rejection at the 5% (respectively 1%) significance level. Values corresponding to models that exhibit the maximum p-value for a given portfolio cardinality are printed in red.

Engle-Colacito regression test for the equally weighted portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	t-stats	-2.42*	-2.48*	-2.32*	-2.41*	-2.54*	**-2.23***
	p-value	0.02	0.01	0.02	0.02	0.01	0.03
10	t-stats	**-2.00***	-2.30*	-2.08*	-2.36*	-2.13*	-2.32*
	p-value	0.05	0.02	0.04	0.02	0.03	0.02
15	t-stats	-1.70	-1.62	-1.89	**-1.60**	-1.71	-1.80
	p-value	0.09	0.11	0.06	0.11	0.09	0.07
20	t-stats	-2.84*	-2.57*	-2.63*	**-2.50***	-2.82*	-3.20*
	p-value	0.00	0.01	0.01	0.01	0.00	0.00
25	t-stats	-2.31*	**-1.87**	-1.98*	-2.03*	-2.25*	-2.63*
	p-value	0.02	0.06	0.05	0.04	0.02	0.01
30	t-stats	-2.97*	-2.82*	-2.87*	**-2.68***	-2.86*	-3.33*
	p-value	0.00	0.00	0.00	0.01	0.00	0.00

Table 4.7: Results of the Engle-Colacito regression test. The t-stat values refer to the intercept λ of the Engle-Colacito regression (4.4) obtained using a HAC estimator. The p-value can be used to test the null hypothesis that $\lambda = 0$. The symbol * (respectively **) indicates rejection at the 5% (respectively 1%) significance level. Values corresponding to models that exhibit the maximum p-value for a given portfolio cardinality are printed in red.

are probably mostly influenced by the common GARCH components of the different models and the differences between the dynamic correlations of the different models are blurred by the common first

MCS of EPA for the minimum variance portfolio squared returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	**2**	**6**	**1**	**3**	**5**	**4**
	p-value	0.562	0.214	1.000	0.562	0.214	0.562
10	Position	**4**	**6**	**3**	**5**	**2**	**1**
	p-value	0.379	0.071	0.379	0.379	0.379	1.000
15	Position	**5**	**3**	**1**	**4**	**2**	**6**
	p-value	0.368	0.676	1.000	0.676	0.676	0.368
20	Position	**4**	**2**	**1**	**5**	**3**	**6**
	p-value	0.339	0.932	1.000	0.339	0.836	0.339
25	Position	**4**	**6**	**1**	**3**	**2**	**5**
	p-value	0.618	0.100	1.000	0.618	0.618	0.618
30	Position	**3**	**6**	**1**	**5**	**2**	**4**
	p-value	0.544	0.054	1.000	0.312	0.808	0.544
	Score	22	29	8	25	16	26

Table 4.8: Model confidence sets based on the predictive ability for squared portfolio returns using the loss function defined in (4.5). See the caption of Table 4.5 for an explanation of the table entries.

MCS of EPA for the equally weighted portfolio squared returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	4	5	2	3	6	**1**
	p-value	0.000	0.000	0.029	0.000	0.000	1.000
10	Position	**1**	4	**2**	6	3	5
	p-value	1.000	0.000	0.438	0.000	0.000	0.000
15	Position	**3**	**2**	6	**1**	**4**	5
	p-value	0.486	0.772	0.001	1.000	0.484	0.003
20	Position	**5**	**2**	**3**	**1**	**4**	6
	p-value	0.061	0.374	0.374	1.000	0.061	0.000
25	Position	5	**1**	**2**	3	4	6
	p-value	0.000	1.000	0.674	0.000	0.000	0.000
30	Position	5	3	4	**1**	2	6
	p-value	0.000	0.036	0.002	1.000	0.036	0.000
	Score	23	17	19	15	23	29

Table 4.9: Results of the model confidence set of EPA for equally weighted portfolio squared returns. The loss function is defined in (4.5). See caption of Table 4.5 for an explanation of the table entries.

HIT test of the minimum variance portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	1%	0.40	0.40	**0.53**	0.40	**0.53**	**0.53**
	p-value	1.0000	**1.0000**	0.9999	1.0000	1.0000	0.9999
	5%	3.60	3.47	3.60	3.60	3.60	**3.74**
	p-value	0.3355	0.2632	0.3367	0.3349	0.3337	**0.5322**
	10%	**8.95**	8.54	8.54	8.54	8.81	8.68
	p-value	**0.5005**	0.2296	0.2278	0.2292	0.4985	0.4310
10	1%	1.60	**1.47**	**1.47**	**1.47**	**1.47**	**1.47**
	p-value	0.9916	0.9924	0.9940	0.9933	**0.9955**	0.9946
	5%	**4.81**	**4.81**	**4.81**	**4.81**	**4.81**	4.67
	p-value	0.7859	0.7784	0.7780	0.7850	0.7894	**0.9189**
	10%	7.34	7.34	**7.74**	7.34	7.48	7.34
	p-value	**0.8484**	0.8423	0.7044	0.7633	0.8345	0.8480
15	1%	**1.07**	1.60	1.34	1.47	1.34	1.20*
	p-value	0.0994	0.1511	**0.3995**	0.1273	0.0579	0.0105
	5%	**4.94**	4.54	4.67	4.81	5.07	4.81
	p-value	**0.8766**	0.6078	0.6573	0.5982	0.8753	0.8643
	10%	8.54	9.08	8.68	**9.21**	8.54	8.81
	p-value	0.9149	**0.9735**	0.9310	0.9350	0.9061	0.9493
20	1%	**1.34***	1.47	**1.34***	**1.34***	**1.34***	**1.34***
	p-value	0.0314	**0.0939**	0.0295	0.0332	0.0305	0.0317
	5%	**4.94**	4.81	**4.94**	4.81	5.07	4.81
	p-value	0.9438	0.9355	0.9262	0.9370	**0.9476**	0.9107
	10%	8.81	10.28	9.61	**10.01**	8.81	8.81
	p-value	0.9216	0.9227	0.9930	**0.9932**	0.9157	0.9228
25	1%	**1.07**	1.60	1.47	1.47	1.20	**1.07**
	p-value	**0.9990**	0.2535	0.5740	0.1508	0.3132	0.9990
	5%	4.67	5.34	**5.07**	4.81	4.81	4.67
	p-value	0.8538	0.3596	**0.9060**	0.4802	0.8694	0.8544
	10%	11.21	12.15	11.62	11.35	11.48	**10.95**
	p-value	0.9361	0.9297	**0.9829**	0.9757	0.9566	0.9548
30	1%	**1.07**	1.60	1.20	1.34	1.20	**1.07**
	p-value	**0.9997**	0.6447	0.2330	0.4011	0.2327	0.9997
	5%	**4.94**	5.87	5.21	5.61	5.07	5.07
	p-value	0.3227	0.2609	0.7368	**0.8300**	0.3966	0.4063
	10%	10.95	12.42	11.75	11.88	11.62	**10.81**
	p-value	0.5710	**0.8997**	0.8530	0.8707	0.7243	0.6226

Table 4.10: Results of the HIT test. For each asset cardinality n and model we report the average number of VaR violations when this risk measure is computed at the 1, 5, and 10% confidence levels. The percentage printed in red corresponds to the model that yields the closest number to the specified confidence level. The p-values correspond to the F-test on the HIT regression (4.6) with five lags. The highest p-values are marked in bold. When they imply that the null hypothesis of independence is rejected at the 5% (respectively 1%) level, the corresponding average number of VaR violations is marked with * (respectively **).

HIT test of the equally weighted portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	1%	0.53	**0.67**	**0.67**	**0.67**	0.53	**0.67**
	p-value	0.9864	0.9986	**0.9987**	0.9985	0.9870	0.9981
	5%	**3.20**	**3.20**	**3.20**	**3.20**	**3.20**	**3.20**
	p-value	0.0982	0.0985	0.0963	0.0986	**0.0995**	0.0975
	10%	**8.28**	8.14	8.14	8.14	8.14	**8.28**
	p-value	0.9162	0.9483	**0.9487**	0.9481	0.9483	0.7887
10	1%	**1.34**	**1.34**	**1.34**	**1.34**	**1.34**	**1.34**
	p-value	**0.9959**	0.9949	0.9948	0.9951	0.9956	0.9944
	5%	**4.01**	**4.01**	**4.01**	**4.01**	**4.01**	**4.01**
	p-value	**0.4566**	0.4355	0.4183	0.4350	0.4543	0.4417
	10%	**6.94**	**6.94**	**6.94**	6.68	**6.94**	6.81
	p-value	0.2370	**0.3799**	0.3659	0.1903	0.2331	0.2884
15	1%	**2.00**	**2.00**	**2.00**	**2.00**	**2.00**	**2.00**
	p-value	0.6435	0.6407	0.6429	0.6388	**0.6465**	0.6439
	5%	4.01	**4.14**	4.01	**4.14**	**4.14**	4.01
	p-value	0.8840	0.9467	0.8897	**0.9521**	0.9361	0.8837
	10%	7.48	7.88	7.74	**8.01**	7.61	7.48
	p-value	0.9053	0.8020	0.8267	0.6327	0.9059	**0.9059**
20	1%	1.34	1.34	1.34	1.47	1.34	**1.20**
	p-value	0.9964	0.9961	0.9957	0.9899	**0.9967**	0.9902
	5%	**3.87**	3.74	**3.87**	3.74	**3.87**	**3.87**
	p-value	0.9354	**0.9414**	0.9368	0.9412	0.9358	0.9351
	10%	7.61	**7.74**	**7.74**	**7.74**	7.61	7.34
	p-value	0.3273	0.3136	0.2648	0.2452	**0.3279**	0.2836
25	1%	**1.07**	1.20	1.34	1.20	1.20	**1.07**
	p-value	0.9959	0.9988	0.9972	0.9985	**0.9989**	0.9952
	5%	4.27	**4.41**	4.14	4.27	4.27	4.14
	p-value	0.6704	0.6125	0.6940	0.6678	0.6694	**0.7031**
	10%	7.08	**7.34**	**7.34**	7.08	7.21	6.81
	p-value	0.6365	0.5347	0.3536	0.6351	0.5319	**0.7141**
30	1%	1.07	1.07	1.07	1.07	1.07	**0.93**
	p-value	0.9796	0.9768	0.9752	0.9844	0.9732	**0.9966**
	5%	3.74	3.74	**3.87**	**3.87**	**3.87**	3.60
	p-value	**0.9449**	0.9423	0.9347	0.9348	0.9348	0.8676
	10%	**6.68**	6.54	6.41	6.54	6.54	6.54
	p-value	0.8804	0.9154	0.8955	**0.9172**	0.9106	0.8993

Table 4.11: Results of the HIT test. For each asset cardinality n and model we report the average number of VaR violations when this risk measure is computed at the 1, 5, and 10% confidence levels. The percentage printed in red corresponds to the model that yields the closest number to the specified confidence level. The p-values correspond to the F-test on the HIT regression (4.6) with five lags. The highest p-values are marked in bold. When they imply that the null hypothesis of independence is rejected at the 5% (respectively 1%) level, the corresponding average number of VaR violations is marked with * (respectively **).

stage models. Therefore, unless the DCC models under consideration produce very different correlation forecasts, it is not surprising that the tests based on portfolio returns do not let us discriminate clearly between the different parameterizations of the correlation dynamics.

On the contrary, the tests behind the MCS based on the correlation of the GARCH standardized returns in (4.3) do not depend directly on the univariate GARCH equations (though they depend on them indirectly, through the degarching of the returns) and, as discussed above, favor clearly the most parsimonious models (Almon, Almon shuffle, scalar, and rank one deficient).

In any case, a reasonable recommendation is that the choice of the optimal model is done *a posteriori* since the grounds for this decision depend on the dataset, time period, and dimensionality of the application of interest. It is clear that all these factors influence the performance of the DCC models and that a careful empirical study has to be conducted in order to select one of them.

5 Conclusions

Several parameterizations for the DCC family of models are used, which go beyond the standard scalar case that appears in most applications. An optimization technique based on Bregman divergences is adapted to effectively perform the two-stage QML estimation of these models that handles all the linear and non-linear parameter constraints that need to be imposed in order to ensure the stationarity of the processes and the positive semidefiniteness of the resulting conditional covariance matrices.

The considered DCC parameterizations are the scalar, Hadamard, and rank deficient ones already proposed in the literature and, additionally, new parameterizations called Almon DCC and Almon shuffle DCC are introduced, which are particular cases of the rank one deficient model. The number of parameters of the latter increases linearly with the dimension of the return process, whereas the corresponding number in the Almon specifications is independent of that dimension. Even though this property also applies to the scalar model, the Almon version is more flexible as it does not impose that the same dynamic pattern applies to all the conditional correlations. All the mathematical and algorithmic details needed to implement the proposed optimization method for each specification of the DCC model are provided (see also the technical appendix). Computer codes will be made available to the community.

An extensive empirical study is carried out using a dataset based on the thirty constituents of the DJIA Index. The proposed estimation procedure is applied to the DCC models under study in various dimensions up to thirty. Moreover, the empirical in- and out-of-sample performances of the different types of parameterizations are compared using a variety of statistical specification tests. The results provide substantial evidence that some non-scalar DCC parameterizations, in particular the Almon (shuffle) and rank one deficient, are worth using in the modeling of the volatility of asset returns in dimensions up to thirty, and perhaps more, though this remains an open issue.

Two extensions are on our research agenda. The first one consists in developing and applying the method of composite quasi-maximum likelihood estimation to the non-scalar DCC models. [PESS14] have found that this method reduces the bias in the estimation of the parameters of the scalar DCC model, so it will be of interest to know whether this result applies also to non-scalar models. The second extension consists in applying the proposed optimization tools to various non-scalar diagonal VEC (DVEC) models and to compare the performance of DCC and DVEC models.

A possible further development is to apply the same tools to the cDCC model of [Aie13]. This extension can be done by keeping the targeting estimator of the constant matrix S as we do for the DCC process. Another, more difficult to implement approach both for DCC and cDCC, will be to estimate S together with the other parameters. An advantage of the cDCC choice is that a profiling method is available in that case, as discussed in [Aie13].

Acknowledgments

We thank the editor and three anonymous referees for their useful and thorough reports that have significantly improved the presentation and the contents of this work. Luc Bauwens acknowledges the support of "Projet d'Actions de Recherche Concertées" 12/17-045 of the "Communauté francaise de Belgique", granted by the "Académie universitaire Louvain". Lyudmila Grigoryeva acknowledges financial support from the Faculty for the Future Program of the Schlumberger Foundation. Lyudmila Grigoryeva and Juan-Pablo Ortega acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493).

References

[Aie13] Gian Piero Aielli. Dynamic Conditional Correlation: on Properties and Estimation. *Journal of Business & Economic Statistics*, 31(3):282–299, July 2013.

[Ale98] Carol Alexander. Orthogonal GARCH. In Carol Alexander, editor, *Mastering Risk*, volume 2, pages 21–38. Financial Times-Prentice Hall, 1998.

[Alm65] S. Almon. The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):178–196, 1965.

[And91] Donald W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.

[BCG06] Monica Billio, Massimiliano Caporin, and Michele Gobbo. Flexible Dynamic Conditional Correlation multivariate GARCH models for asset allocation. *Applied Financial Economics Letters*, 2(2):123–130, 2006.

[Bol86] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[Bol90] Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, 72(3):498–505, 1990.

[BR97] R. B. Bapat and T. E. S. Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, 1997.

[Bre67] L.M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *Zh. vychisl. mat. Mat. Fiz.*, 7(3):620–631, 1967.

[CES06] Lorenzo Cappiello, Robert F. Engle, and Kevin K. Sheppard. Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Economics*, 4(4):537–572, 2006.

[CM12] Massimiliano Caporin and Michael McAleer. Do we really need both BEKK and DCC? A tale of two multivariate GARCH models. *Journal of Economic Surveys*, 26(4):736–751, September 2012.

[CO14] Stéphane Chrétien and Juan-Pablo Ortega. Multivariate GARCH estimation via a Bregman-proximal trust-region method. *Computational Statistics and Data Analysis*, 76:210–236, 2014.

[DT07] Inderjit S Dhillon and Joel A Tropp. Matrix nearness problems with Bregman divergences. *SIAM J. Matrix Anal. Appl.*, 29(4):1120–1146, 2007.

[EC06] Robert F. Engle and Riccardo Colacito. Testing and Valuing Dynamic Correlations for Asset Allocation. *Journal of Business & Economic Statistics*, 24(2):238–253, 2006.

[EM04] Robert F Engle and Simone Manganelli. Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381, October 2004.

[Eng02] Robert F Engle. Dynamic conditional correlation -a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, 20:339–350, 2002.

[Eng08] Robert F. Engle. High dimensional dynamic correlations. In J. L. Castle and N. Shephard, editors, *The Methodology and Practice of Econometrics: Papers in Honour of David F Hendry*. Oxford University Press, Oxford, 2008.

[Eng09] Robert Engle. *Anticipating Correlations*. Princeton University Press, Princeton, NJ, 2009.

[HF09] Christian M. Hafner and P. H. Franses. A generalized Dynamic Conditional Correlation model: simulation and application to many assets. *Econometric Reviews*, 28(6):612–631, 2009.

[HLN03] Peter Reinhard Hansen, Asger Lunde, and James M. Nason. Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics*, 65(s1):839–861, December 2003.

[HLN11] Peter Reinhard Hansen, Asger Lunde, and James M. Nason. The model confidence set. *Econometrica*, 79(2):453–497, 2011.

[JS61] W James and Charles Stein. Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 361–379. Univ. California Press, Berkeley, Calif., 1961.

[KSD09a] Brian Kulis, Mátyás A Sustik, and Inderjit S Dhillon. Low-rank kernel learning with Bregman matrix divergences. *J. Mach. Learn. Res.*, 10:341–376, 2009.

[KSD09b] Brian Kulis, Suvrit S Sustik, and Inderjit S Dhillon. Convex perturbations for scalable semidefinite programming. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS) 2009*, pages 296–303, 2009.

[NSS14] Diaa Noureldin, Neil Shephard, and Kevin Sheppard. Multivariate Rotated ARCH models. *Journal of Econometrics*, 179(1):16–30, 2014.

[PESS14] Cavit Pakel, Robert F. Engle, Neil Shephard, and Kevin K. Sheppard. Fitting Vast Dimensional Time-Varying Covariance Models, 2014.

[TT02] Y. K. Tse and A. K. C. Tsui. A multivariate GARCH with time-varying correlations. *Journal of Business and Economic Statistics*, 20:351–362, 2002.

Technical appendix to the paper "Estimation and empirical performance of non-scalar dynamic conditional correlation models"

Luc Bauwens[1], Lyudmila Grigoryeva[2], and Juan-Pablo Ortega[3]

Abstract

This document contains complementary material to the paper [BGO14]. Sections in this appendix are indexed by letters and propositions and formulas by a letter followed by a number (e.g. A.1). Sections, propositions, and formulas in the paper are referenced by numbers. Section A contains the notation that is used throughout this technical appendix and various background results that are quoted in the sequel. Sections B and C provide the proofs of Propositions 2.1 and 2.2, respectively. Section D presents the computation of the gradient of the log-likelihood function. Section E presents the results needed for implementing the constrained optimization algorithm to the DCC models. Section F contains the detailed derivations of these results. Section G provides the tables containing the empirical results for a second dataset.

[1]Université catholique de Louvain, CORE, Voie du Roman Pays 34 L1.03.01, B-1348 Louvain-La-Neuve. Belgium; University of Johannesburg, Department of Economics, Johannesburg, South Africa. luc.bauwens@uclouvain.be

[2]Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, UFR des Sciences et Techniques. 16, route de Gray. F-25030 Besançon cedex. France. Lyudmyla.Grygoryeva@univ-fcomte.fr

[3]Centre National de la Recherche Scientifique, Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, UFR des Sciences et Techniques. 16, route de Gray. F-25030 Besançon cedex. France. Juan-Pablo.Ortega@univ-fcomte.fr

Acknowledgments: Lyudmila Grigoryeva and Juan-Pablo Ortega acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493). Lyudmila Grigoryeva acknowledges financial support from the Faculty for the Future Program of the Schlumberger Foundation. Luc Bauwens acknowledges the support of "Projet d'Actions de Recherche Concertées" 12/17-045 of the "Communauté francaise de Belgique", granted by the "Académie universitaire Louvain".

A Notation and preliminaries

A.1 Vectors and matrices

Vector notation: a column vector is denoted by a bold lower case symbol like $\mathbf{r}$ and $\mathbf{r}^\top$ indicates its transpose. Given a vector $\mathbf{v} \in \mathbb{R}^n$, we denote its entries by v_i, with $i \in \{1, \dots, n\}$; we also write $\mathbf{v} = (v_i)_{i\in\{1,\dots,n\}}$. The symbols $\mathbf{i}_n, \mathbf{0}_n \in \mathbb{R}^n$ stand for the vectors of length n consisting of ones and of zeros, respectively. Additionally, given $n \in \mathbb{N}$, we define the vectors $\mathbf{k}_n^1 := (1, 2, \dots, n)^\top$, $\mathbf{k}_n^2 := \left(1, 2^2, \dots, n^2\right)^\top \in \mathbb{R}^n$; $\mathbf{e}_n^{(i)} \in \mathbb{R}^n$, $i \in \{1, \dots, n\}$ denotes the canonical unit vector of length n determined by $\mathbf{e}_n^{(i)} = (\delta_{ij})_{j\in\{1,\dots,n\}}$.

Matrix notation: we denote by $\mathbb{M}_{n,m}$ the space of real $n \times m$ matrices with $m, n \in \mathbb{N}$. When $n = m$, we use the symbols $\mathbb{M}_n$ and $\mathbb{D}_n$ to refer to the space of square and diagonal matrices of order n, respectively. Given a matrix $A \in \mathbb{M}_{n,m}$, we denote its components by A_{ij} and we write $A = (A_{ij})$, with $i \in \{1, \dots, n\}$, $j \in \{1, \dots m\}$. The symbol $\mathbb{L}_{n,m}$ denotes the subspace of lower triangular matrices, that is, matrices that have zeros above the main diagonal:

$$\mathbb{L}_{n,m} = \{A \in \mathbb{M}_{n,m} \mid A_{ij} = 0, j > i\} \subset \mathbb{M}_{n,m}.$$

We denote by $\mathbb{L}_{n,m}^+ \subset \mathbb{L}_{n,m}$ (respectively $\mathbb{L}_{n,m}^- \subset \mathbb{L}_{n,m}$) the cone of matrices in $\mathbb{L}_{n,m}$ whose elements in the main diagonal are all positive (respectively negative). We use $\mathbb{S}_n$ to denote the subspace $\mathbb{S}_n \subset \mathbb{M}_n$ of symmetric matrices:

$$\mathbb{S}_n = \left\{A \in \mathbb{M}_n \mid A^\top = A\right\},$$

and we use $\mathbb{S}_n^+$ (respectively $\mathbb{S}_n^-$) to refer to the cone $\mathbb{S}_n^+ \subset \mathbb{S}_n$ (respectively $\mathbb{S}_n^- \subset \mathbb{S}_n$) of positive (respectively negative) semidefinite matrices. We write $A \succeq 0$ (respectively $A \preceq 0$) when $A \in \mathbb{S}_n^+$ (respectively $A \in \mathbb{S}_n^-$). The symbol $\mathbb{I}_n \in \mathbb{D}_n$ denotes the identity matrix and $\mathbb{O}_n \subset \mathbb{M}_n$ is the subspace of orthogonal matrices, that is, $\mathbb{O}_n := \{A \in \mathbb{M}_n \mid AA^\top = \mathbb{I}_n\}$.

The Frobenius inner product is defined on the space $\mathbb{M}_{n,m}$ as:

$$\langle A, B\rangle := \operatorname{tr}\left(AB^\top\right) = \operatorname{tr}\left(A^\top B\right), \quad A, B \in \mathbb{M}_{n,m}. \tag{A.1}$$

The symbol tr denotes the trace of the matrix. This inner product induces the Frobenius norm that we denote as $\|A\| := \langle A, A\rangle^{1/2}$. Given a linear operator $\mathcal{A} : \mathbb{M}_{n,m} \longrightarrow \mathbb{M}_{p,q}$, we denote by $\mathcal{A}^* : \mathbb{M}_{p,q} \longrightarrow \mathbb{M}_{n,m}$ its adjoint operator with respect to (A.1) by the relation

$$\langle B, \mathcal{A}(C)\rangle = \langle \mathcal{A}^*(B), C\rangle, \quad \text{for any} \quad B \in \mathbb{M}_{n,m}, \text{ and } C \in \mathbb{M}_{p,q}.$$

The Hadamard product of matrices: given two matrices $A, B \in \mathbb{M}_{n,m}$, we denote by $A \odot B \in \mathbb{M}_{n,m}$ their elementwise multiplication matrix or Hadamard product, that is:

$$(A \odot B)_{ij} := A_{ij}B_{ij} \text{ for all } i \in \{1, \dots, n\}, \ j \in \{1, \dots, m\}. \tag{A.2}$$

The main properties of the Hadamard product that are used in the sequel are the following:

(i) ***The Hadamard product of two vectors:*** given two arbitrary vectors $\mathbf{u}, \mathbf{w} \in \mathbb{R}^n$, the following relation holds true

$$\mathbf{u} \odot \mathbf{w} = U\mathbf{w}, \tag{A.3}$$

where $U \in \mathbb{D}_n$ is defined by $U_{ii} := u_i$, for all $i \in \{1, \dots, n\}$, that is, $U := \operatorname{diag}(\mathbf{u})$, where the operator diag is defined in the following subsection.

(ii) ***The Hadamard product trace property:*** consider the matrices $A, B, C \in \mathbb{M}_{n,m}$. Then the following relation holds (see for instance [HJ94, page 304])

$$((A \odot B)C^\top)_{ii} = ((A \odot C)B^\top)_{ii} \text{ for all } i \in \{1, \dots, n\}.$$

This leads to the equality

$$\operatorname{tr}\left((A \odot B)\, C^\top\right) = \operatorname{tr}\left((A \odot C)\, B^\top\right), \tag{A.4}$$

which we refer to as *the Hadamard product trace property.*

(iii) ***Schur Product Theorem***: let $A, B \in \mathbb{M}_n$ be positive semidefinite matrices. Then $A \odot B$ is also positive semidefinite. See [BR97] for a proof.

A.2 Operators and their adjoints

We recall some standard matrix operators and introduce several new ones that we use in the following sections.

The diag and Diag operators: we denote as Diag the operator $\operatorname{Diag} : \mathbb{M}_n \longrightarrow \mathbb{D}_n$ that sets equal to zero all the components of a square matrix except for those that are on the main diagonal. The operator $\operatorname{diag} : \mathbb{R}^n \longrightarrow \mathbb{D}_n$ takes a given vector and constructs a diagonal matrix with its entries in the main diagonal. We denote by $\operatorname{diag}^{-1} : \mathbb{D}_n \longrightarrow \mathbb{R}^n$ the inverse of the diag operator. The adjoint operator of Diag (respectively diag) is denoted by $\operatorname{Diag}^* : \mathbb{D}_n \longrightarrow \mathbb{M}_n$ (respectively $\operatorname{diag}^* : \mathbb{D}_n \longrightarrow \mathbb{R}^n$); it is easy to see that Diag^* is just the injection $\mathbb{D}_n \hookrightarrow \mathbb{M}_n$ and that $\operatorname{diag}^* = \operatorname{diag}^{-1}$.

The vec and mat operators: Given a matrix $A \in \mathbb{M}_{n,m}$, we denote by vec the operator that transforms A into a vector of length nm by stacking all its columns, namely,

$$\operatorname{vec} : \mathbb{M}_{n,m} \longrightarrow \mathbb{R}^{nm}, \quad \operatorname{vec}(A) = (A_{11}, \dots, A_{n1}, \dots, A_{1m}, \dots, A_{nm})^\top.$$

The inverse of this operator is denoted as $\operatorname{mat} : \mathbb{R}^{nm} \longrightarrow \mathbb{M}_{n,m}$.

The vech and math operators: we denote by vech the operator that stacks the elements on and below the main diagonal of a symmetric matrix into a vector of length $N := \frac{1}{2}n(n+1)$, that is,

$$\operatorname{vech} : \mathbb{S}_n \longrightarrow \mathbb{R}^N, \quad \operatorname{vech}(A) = (A_{11}, \dots, A_{n1}, A_{22}, \dots, A_{n2}, \dots, A_{nn})^\top, A \in \mathbb{S}_n,$$

and we denote the inverse of this operator by $\operatorname{math} : \mathbb{R}^N \longrightarrow \mathbb{S}_n$.
The adjoint map of vech (respectively math) is denoted by $\operatorname{vech}^* : \mathbb{R}^N \longrightarrow \mathbb{S}_n$ (respectively $\operatorname{math}^* : \mathbb{S}_n \longrightarrow \mathbb{R}^N$). In [CO14] it is shown that given $A \in \mathbb{S}_n$ and $\mathbf{v} \in \mathbb{R}^N$, the following relations hold true:

$$\operatorname{vech}^*(\mathbf{v}) = \frac{1}{2}\left(\operatorname{math}(\mathbf{v}) + \operatorname{Diag}(\operatorname{math}(\mathbf{v}))\right), \tag{A.5}$$

$$\operatorname{math}^*(A) = 2\ \operatorname{vech}(A - \frac{1}{2}\ \operatorname{Diag}(A)). \tag{A.6}$$

The relation between the vech, math, vec, and Diag operators: given a matrix $A \in \mathbb{S}_n$ and $N := \frac{1}{2}n(n+1)$, we denote by $L_n \in \mathbb{M}_{N,n^2}$ and by $D_n \in \mathbb{M}_{n^2,N}$ the elimination and the duplication matrices, respectively (see [Lü05]). These matrices satisfy:

$$\operatorname{vech}(A) = L_n \operatorname{vec}(A), \tag{A.7}$$

$$\operatorname{vec}(A) = D_n \operatorname{vech}(A). \tag{A.8}$$

Given $A \in \mathbb{M}_n$, we define the diagonalization matrix $P_n^d \in \mathbb{M}_n$ via the relation:

$$\text{math}\left(P_n^d \text{vech}\left(A\right)\right) = \text{Diag}\left(A\right). \tag{A.9}$$

A well-known property of the vec operator that is exploited in the following sections is

$$\text{vec}\left(ABC\right) = \left(C^\top \otimes A\right)\text{vec}(B). \tag{A.10}$$

The mat_r and vec_r operators: let $r \leq n \in \mathbb{N}$, $N^* := nr - \frac{1}{2}r\left(r-1\right)$ and define the operator $\text{mat}_r : \mathbb{R}^{N^*} \longrightarrow \mathbb{L}_{n,r}$ that transforms a vector of length N^* into the lower triangular $n \times r$ matrix defined by,

$$\text{mat}_r\left(\mathbf{v}\right) = \begin{pmatrix} v_1 & 0 & \cdots & 0 \\ v_2 & v_{n+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ v_r & v_{n+r-1} & \cdots & v_{N^*-n+r} \\ \vdots & \vdots & \ddots & \vdots \\ v_n & v_{2n-1} & \cdots & v_{N^*} \end{pmatrix}, \quad \text{for any } \mathbf{v} \in \mathbb{R}^{N^*}. \tag{A.11}$$

We denote the inverse of this operator as $\text{vec}_r : \mathbb{L}_{n,r} \longrightarrow \mathbb{R}^{N^*}$. We note for future reference that the elements in the main diagonal of (A.11) are given by

$$\{v_{i_1}, \ldots, v_{i_r}\} \quad \text{with} \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1, \ldots, r\}. \tag{A.12}$$

The following proposition characterizes the adjoint maps of mat_r and vec_r, respectively. Its proof is provided in Section A.3.

Proposition A.1 *Given $r \leq n \in \mathbb{N}$ and $N^* = nr - \frac{1}{2}r\left(r-1\right)$, let $A \in \mathbb{L}_{n,r}$ and $\mathbf{v} \in \mathbb{R}^{N^*}$ arbitrary. Let $\text{mat}_r^* : \mathbb{L}_{n,r} \longrightarrow \mathbb{R}^{N^*}$ and $\text{vec}_r^* : \mathbb{R}^{N^*} \longrightarrow \mathbb{L}_{n,r}$ be the adjoint maps of mat_r and vec_r, respectively. Then, the following relations hold true:*

$$\text{mat}_r^*\left(A\right) = \text{vec}_r\left(A\right), \tag{A.13}$$

$$\text{vec}_r^*\left(\mathbf{v}\right) = \text{mat}_r\left(\mathbf{v}\right). \tag{A.14}$$

The Almon lag operator and its tangent map: using the Almon lag function [Alm65], we define the Almon lag operator $\text{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$, with $n \in \mathbb{N}$ and $\mathbf{v} \in \mathbb{R}^3$, as

$$\left(\text{alm}_n(\mathbf{v})\right)_i := v_1 + \exp(v_2 i + v_3 i^2), \text{ for } i \in \{1, \ldots, n\}. \tag{A.15}$$

The tangent map $T_{\mathbf{v}}\text{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$ is determined by the equality

$$T_{\mathbf{v}}\text{alm}_n \cdot \boldsymbol{\delta}\mathbf{v} = K_{\mathbf{v}} \cdot \boldsymbol{\delta}\mathbf{v}, \text{ with } K_{\mathbf{v}} := \left(\mathbf{i}_n \mid \mathbf{k}_n^1 \odot \text{alm}_n(\bar{\mathbf{v}}) \mid \mathbf{k}_n^2 \odot \text{alm}_n(\bar{\mathbf{v}})\right) \in \mathbb{M}_{n,3}, \boldsymbol{\delta}\mathbf{v} \in \mathbb{R}^3, \tag{A.16}$$

where the symbol $\mid$ denotes vertical concatenation of matrices (or vectors), the vectors $\mathbf{i}_n, \mathbf{k}_n^1, \mathbf{k}_n^2 \in \mathbb{R}^n$ were introduced in Subsection A.1, and $\bar{\mathbf{v}} \in \mathbb{R}^3$ is obtained out of the vector $\mathbf{v} \in \mathbb{R}^3$ by setting its first component equal to zero, namely, $\bar{\mathbf{v}} := (0, v_2, v_3)^\top$.
The adjoint $T_{\mathbf{v}}^*\text{alm}_n : \mathbb{R}^n \longrightarrow \mathbb{R}^3$ of the tangent map $T_{\mathbf{v}}\text{alm}_n$ is determined by the relation

$$T_{\mathbf{v}}^*\text{alm}_n(\mathbf{u}) = K_{\mathbf{v}}^\top \cdot \mathbf{u}, \quad \text{for any } \mathbf{u} \in \mathbb{R}^n. \tag{A.17}$$

A.3 Proof of Proposition A.1

In order to prove the Proposition A.1, we introduce the auxiliary operator σ.
The operator σ: Let $r \leq n \in \mathbb{N}$, $N^* = nr - \frac{1}{2}r\,(r-1)$, and

$$S = \{(i,j)\,|i \in \{1,\dots,n\}\,, j \in \{1,\dots,r\}\,, i \geq j\}\,.$$

Given a matrix $A \in \mathbb{L}_{n,r}$ and $\mathbf{v} = \mathrm{vec}_r\,(A) \in \mathbb{R}^{N^*}$ the operator $\sigma : S \longrightarrow \{1,\dots,N^*\}$ assigns to the position of the entry (i,j), $i \geq j$, of the matrix A the position of the corresponding element of $\mathbf{v}$ in the vec_r representation. We refer to the inverse of this operator as $\sigma^{-1} : \{1,\dots,N^*\} \longrightarrow S$.

Proof of Proposition A.1: In order to prove (A.13), we use the following chain of equalities:

$$\begin{aligned}
\langle A, \mathrm{mat}_r\,(\mathbf{v})\rangle &= \mathrm{tr}\left(A^\top \cdot \mathrm{mat}_r\,(\mathbf{v})\right) = \sum_{i=1}^{n}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} = \sum_{i=r}^{n}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} \\
&+ \sum_{i=1}^{r-1}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} = \sum_{i=r}^{n}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} + \sum_{i=1}^{r-1}\sum_{j=1}^{i} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} \\
&+ \sum_{i=1}^{r-1}\sum_{j=i+1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} = \sum_{i=r}^{n}\sum_{j=1}^{r} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} + \sum_{i=1}^{r-1}\sum_{j=1}^{i} A_{ij}(\mathrm{mat}_r\,(\mathbf{v}))_{ij} \\
&= \sum_{i=r}^{n}\sum_{j=1}^{r} A_{ij}v_{\sigma(i,j)} + \sum_{i=1}^{r-1}\sum_{j=1}^{i} A_{ij}v_{\sigma(i,j)} = \sum_{q=1}^{N^*} A_{\sigma^{-1}(q)}v_q = \langle \mathrm{vec}_r\,(A)\,, \mathbf{v}\rangle.
\end{aligned}$$

Since $A \in \mathbb{L}_{n,r}$ and $\mathbf{v} \in \mathbb{R}^{N^*}$ in these equalities are arbitrary, the identity $\langle A, \mathrm{mat}_r\,(\mathbf{v})\rangle = \langle \mathrm{vec}_r\,(A)\,, \mathbf{v}\rangle$ ensures that $\langle \mathrm{mat}_r^*\,(A)\,, \mathbf{v}\rangle = \langle \mathrm{vec}_r\,(A)\,, \mathbf{v}\rangle$, which yields (A.13). In order to prove the relation (A.14) we write down:

$$\begin{aligned}
\langle \mathbf{v}, \mathrm{vec}_r\,(A)\rangle &= \mathrm{tr}\left(\mathbf{v}^\top \mathrm{vec}_r\,(A)\right) = \sum_{i=1}^{N^*} v_i(\mathrm{vec}_r\,(A))_i = \sum_{i=1}^{n}\sum_{j=1}^{r} v_{\sigma(i,j)}(\mathrm{vec}_r\,(A))_{\sigma(i,j)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{r}(\mathrm{mat}_r\,(\mathbf{v}))_{ij}A_{ij} = \mathrm{tr}((\mathrm{mat}_r(\mathbf{v})^\top \cdot A) = \langle \mathrm{mat}_r\,(\mathbf{v})\,, A\rangle,
\end{aligned}$$

which yields $\langle \mathrm{vec}_r^*\,(\mathbf{v})\,, A\rangle = \langle \mathrm{mat}_r\,(\mathbf{v})\,, A\rangle$ and hence proves (A.14).

B Proof of Proposition 2.1

We start by recalling the general fact that any positive definite matrix $M \in \mathbb{S}_n^+$ has positive diagonal entries, that is,

$$M_{ii} > 0, \qquad \text{for any } i \in \{1,\dots,n\}. \tag{B.1}$$

Indeed, given a canonical vector $\mathbf{e}_n^{(i)}$, $i \in \{1,\dots,n\}$, $M_{ii} = \langle \mathbf{e}_n^{(i)}, M\mathbf{e}_n^{(i)}\rangle$, which is positive by the assumption $M \in \mathbb{S}_n^+$. Let now $A, B \in \mathbb{S}_n$ be two parameter matrices that satisfy the constraints (2.10), (2.11) and define $C := \left(\mathbf{i}_n\mathbf{i}_n^\top - A - B\right) \odot S$. Notice first that the condition (2.11) implies by (B.1) that

$$C_{ii} = (1 - A_{ii} - B_{ii})S_{ii} > 0, \qquad \text{for any } i \in \{1,\dots,n\}. \tag{B.2}$$

Additionally, since the matrix S is approximated by the empirical covariance (2.9) then $S \in \mathbb{S}_n^+$ and hence, again by (B.1), we can state that $S_{ii} > 0$ almost surely, for all $i \in \{1,\dots,n\}$. This automatically yields that

$$A_{ii} + B_{ii} < 1, \tag{B.3}$$

with $A_{ii}, B_{ii} \geq 0$, $i \in \{1, \dots, n\}$, by (2.10) and (B.1). We now easily prove that $|A_{ij} + B_{ij}| < 1$ for any $i, j \in \{1, \dots, n\}$ as a corollary of (B.3) and the Cauchy-Bunyakovsky-Schwarz inequality. Indeed, since $A, B \in \mathbb{S}_n^+$, then the sum $A + B \in \mathbb{S}_n^+$ and hence we can write the following relations for all pairs of canonical vectors $\mathbf{e}_n^{(i)}$ and $\mathbf{e}_n^{(j)}$, $i, j \in \{1, \dots, n\}$,

$$\begin{aligned} |A_{ij} + B_{ij}| =& |\mathbf{e}_n^{(i)\top}(A+B)\mathbf{e}_n^{(j)}| = |\mathbf{e}_n^{(i)\top}(A+B)^{1/2}(A+B)^{1/2}\,\mathbf{e}_n^{(j)}| = |\langle (A+B)^{1/2}\,\mathbf{e}_n^{(i)}, (A+B)^{1/2}\,\mathbf{e}_n^{(j)}\rangle| \\ \leq& \sqrt{\langle (A+B)^{1/2}\,\mathbf{e}_n^{(i)}, (A+B)^{1/2}\,\mathbf{e}_n^{(i)}\rangle \langle (A+B)^{1/2}\,\mathbf{e}_n^{(j)}, (A+B)^{1/2}\,\mathbf{e}_n^{(j)}\rangle} \\ =& \sqrt{(A_{ii}+B_{ii})(A_{jj}+B_{jj})} < 1, \end{aligned} \tag{B.4}$$

where in the last inequality we used that A, $B \in \mathbb{S}_n^+$ satisfy (B.3). The relations (B.3) and (B.4) automatically yield (2.6), as required.

C Proof of Proposition 2.2

We start by showing that the map Ψ is injective. Let $A, B \in \mathbb{L}_{n,m}^+$ be such that $[A] = [B]$. This implies the existence of an element $O \in \mathbb{O}(r)$ such that $AO^{-1} = B$ or, equivalently, $OA^\top = B^\top$. This equality can be written in terms of matrix entries as

$$O \left(\begin{array}{ccccc|c} a_{11} & a_{21} & a_{31} & \cdots & a_{r1} & \\ 0 & a_{22} & a_{32} & \cdots & a_{r2} & \\ 0 & 0 & a_{33} & \cdots & a_{r3} & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & a_{rr} & \end{array} \middle| \, \widetilde{A} \right) = \left(\begin{array}{ccccc|c} b_{11} & b_{21} & b_{31} & \cdots & b_{r1} & \\ 0 & b_{22} & b_{32} & \cdots & b_{r2} & \\ 0 & 0 & b_{33} & \cdots & b_{r3} & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & b_{rr} & \end{array} \middle| \, \widetilde{B} \right), \tag{C.1}$$

with $\widetilde{A}, \widetilde{B} \in \mathbb{M}_{r,n-r}$. We now proceed recursively by analyzing the n different equations included in (C.1). First of all, since $A, B \in \mathbb{L}_{n,m}^+$, then $a_{11}, b_{11} > 0$ and hence the equation $O\,(a_{11}, 0, \cdots, 0)^\top = (b_{11}, 0, \dots, 0)^\top$ implies that $a_{11} = b_{11}$ and that $O \in \mathbb{O}(r)$ belongs to the subgroup of $\mathbb{O}(r)$ isomorphic to $\mathbb{O}(r-1)$ that leaves invariant the vectors in $\text{span}\{\mathbf{e}_r^{(1)}\}$. This statement, together with the second equation included in (C.1), that is, $O\,(a_{21}, a_{22}, 0, \dots, 0)^\top = (b_{21}, b_{22}, 0, \dots, 0)^\top$ imply that $a_{21} = b_{21}$ and that $O\,(0, a_{22}, 0, \dots, 0)^\top = (0, b_{22}, 0, \dots, 0)^\top$. Since $a_{22}, b_{22} > 0$ then $a_{22} = b_{22}$ necessarily and we can conclude that $O \in \mathbb{O}(r)$ belongs to the subgroup of $\mathbb{O}(r)$ isomorphic to $\mathbb{O}(r-2)$ that leaves invariant the vectors in $\text{span}\{\mathbf{e}_r^{(1)}, \mathbf{e}_r^{(2)}\}$. We are able to conclude by repeating this procedure r times that $O = \mathbb{I}_r$ necessarily and that $A = B$, as required.

In order to show that Ψ is also surjective, we have to prove that for any $[B] \in \mathbb{M}_{n,r}/\mathbb{O}(r)$, there exists $A \in \mathbb{L}_{n,m}^+$ such that $\Psi(A) = [B]$. Let $B \in \mathbb{M}_{n,r}$ be an arbitrary element in the orbit $[B]$ such that

$$B^\top = \left(\begin{array}{ccccc|c} b_{11} & b_{21} & b_{31} & \cdots & b_{r1} & \\ b_{12} & b_{22} & b_{32} & \cdots & b_{r2} & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ b_{1r} & b_{2r} & b_{3r} & \cdots & b_{rr} & \end{array} \middle| \, \widetilde{B} \right),$$

with $\widetilde{B} \in \mathbb{M}_{r,n-r}$. Let $O_1 \in \mathbb{O}(r)$ be such that $O_1\,(b_{11}, b_{12}, \dots, b_{1r})^\top = \left(b_{11}^1, 0, \dots, 0\right)^\top$, for some $b_{11}^1 > 0$ and let

$$\left(\begin{array}{ccccc|c} b_{11}^1 & b_{21}^1 & b_{31}^1 & \cdots & b_{r1}^1 & \\ 0 & b_{22}^1 & b_{32}^1 & \cdots & b_{r2}^1 & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & b_{2r}^1 & b_{3r}^1 & \cdots & b_{rr}^1 & \end{array} \middle| \, \widetilde{B}^1 \right) := O_1 B^\top.$$

Consider now another element $O_2 \in \mathbb{O}(r)$ that leaves invariant the vectors in $\text{span}\{\mathbf{e}_r^{(1)}\}$ and such that $O_2 \left(b_{21}^1, b_{22}^1, \ldots, b_{2r}^1\right)^\top = \left(b_{21}^1, b_{22}^2, 0, \ldots, 0\right)^\top$, for some $b_{22}^2 > 0$, and let

$$\left(\begin{array}{ccccc|c} b_{11}^1 & b_{21}^1 & b_{31}^1 & \cdots & b_{r1}^1 & \\ 0 & b_{22}^2 & b_{32}^2 & \cdots & b_{r2}^2 & \\ 0 & 0 & b_{33}^2 & \cdots & b_{r3}^2 & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \widetilde{B}^2 \\ 0 & 0 & b_{3r}^2 & \cdots & b_{rr}^2 & \end{array}\right) := O_2 O_1 B^\top.$$

If we iterate r times this construction we obtain r elements $O_1, O_2, \ldots, O_r \in \mathbb{O}(r)$ such that the matrix A defined by $A^\top := O_r O_{r-1} \cdots O_2 O_1 B^\top$ belongs to $\mathbb{L}_{n,m}^+$ and since by construction $[A] = [B]$ we have that $\Psi(A) = [A] = [B]$, and the result follows.

D Gradient of log-likelihood function

D.1 Direct derivations

Proposition D.1 *Let $\mathbf{r} = \{\mathbf{r}_1, \ldots, \mathbf{r}_T\}$ be a sample with $\mathbf{r}_t \in \mathbb{R}^n$, $t \in \{1, \ldots, T\}$. Let $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathcal{P} \times \mathcal{P}$, $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (A(\boldsymbol{\theta}_1), B(\boldsymbol{\theta}_2)) \in \mathbb{S}_n \times \mathbb{S}_n$, and let $\log L(\boldsymbol{\theta}; \mathbf{r})$ be the log-likelihood in (3.1)-(3.2). Then,*

$$\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}; \mathbf{r}) = \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} l_t(\boldsymbol{\theta}; \mathbf{r}_t) = \sum_{t=1}^{T} T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta} \cdot T_{\boldsymbol{\Theta}}^* H_t \cdot \nabla_{H_t} l_t(\boldsymbol{\theta}; \mathbf{r}_t), \tag{D.1}$$

with

$$\nabla_{H_t} l_t(\boldsymbol{\theta}; \mathbf{r}_t) = -\frac{1}{2}\left[H_t^{-1} - H_t^{-1} \mathbf{r}_t \mathbf{r}_t^\top H_t^{-1}\right]. \tag{D.2}$$

In the relation (D.1), the differential operator $T_{\boldsymbol{\Theta}}^* H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is the adjoint of the map $T_{\boldsymbol{\Theta}} H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$. For each component Θ (that is $A(\boldsymbol{\theta}_1)$ and $B(\boldsymbol{\theta}_2)$) of $\boldsymbol{\Theta}$ and for any $\Delta \in \mathbb{S}_n$, $T_{\boldsymbol{\Theta}}^* H_t$ is determined by the expression:

$$T_\Theta^* H_t \cdot \Delta = T_\Theta^* Q_t \left[Q_t^{*-1} D_t \Delta D_t Q_t^{*-1} - \frac{1}{2}\text{Diag}(Q_t^{*-2}(D_t \Delta D_t Q_t^{*-1} Q_t + Q_t Q_t^{*-1} D_t \Delta D_t) Q_t^{*-1})\right]. \tag{D.3}$$

Additionally, the differential operator $T_{\boldsymbol{\Theta}}^* Q_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is the adjoint of the map $T_{\boldsymbol{\Theta}} Q_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$; for each component $A(\boldsymbol{\theta}_1)$ and $B(\boldsymbol{\theta}_2)$ of $\boldsymbol{\Theta}$ and for any $\Delta \in \mathbb{S}_n$, $T_{\boldsymbol{\Theta}}^* Q_t$ is determined by the recursions:

$$T_A^* Q_t \cdot \Delta = \Delta \odot \left(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}^\top - S\right) + T_A^* Q_{t-1}[\Delta \odot B], \tag{D.4}$$

$$T_B^* Q_t \cdot \Delta = \Delta \odot (Q_{t-1} - S) + T_B^* Q_{t-1}[\Delta \odot B], \tag{D.5}$$

that are initialized by setting $T_A^* Q_0 = 0$ and $T_B^* Q_0 = 0$.
Finally, the differential operator $T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta} : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathcal{P} \times \mathcal{P}$ is the adjoint of the map $T_{\boldsymbol{\theta}} \boldsymbol{\Theta} : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$, with $\mathcal{P} \times \mathcal{P}$ the intrinsic $\boldsymbol{\theta}$ parameter space and parameterization $\boldsymbol{\Theta}(\boldsymbol{\theta})$ associated to each of the model subfamilies considered in Section 2. For a given pair $\Delta_1, \Delta_2 \in \mathbb{S}_n$ these maps are determined by the following expressions:

(i) ***The Hadamard DCC family:*** *let $n \in \mathbb{N}$, $N := \frac{1}{2} n(n+1)$. In this case, the intrinsic parameter subspace $\mathcal{P}$ is $\mathbb{R}^N$, $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b})$, and $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (\text{math}(\boldsymbol{a}), \text{math}(\boldsymbol{b}))$, for any $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$. Moreover,*

$$\begin{array}{rccc} T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta}: & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R}^N \times \mathbb{R}^N \\ & (\Delta_1, \Delta_2) & \longmapsto & (\text{math}^*(\Delta_1), \text{math}^*(\Delta_2)). \end{array} \tag{D.6}$$

(ii) ***The rank deficient DCC family with rank*** r***:*** *let* $r < n \in \mathbb{N}$, $N^* := nr - \frac{1}{2}r\,(r-1)$. *In this case the intrinsic parameter subspace* $\mathcal{P}$ *is* $\mathbb{R}^{N^*}$, $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b})$, *and*

$$\boldsymbol{\Theta}(\boldsymbol{\theta}) := (\mathrm{mat}_r(\boldsymbol{a})(\mathrm{mat}_r(\boldsymbol{a}))^\top, \mathrm{mat}_r(\boldsymbol{b})(\mathrm{mat}_r(\boldsymbol{b}))^\top), \quad \textit{for any} \quad \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{N^*}.$$

Moreover,

$$\begin{array}{rccc} T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}: & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \\ & (\Delta_1, \Delta_2) & \longmapsto & 2\,(\mathrm{vec}_r(\Delta_1 \mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(\Delta_2 \mathrm{mat}_r(\boldsymbol{b}))). \end{array} \tag{D.7}$$

(iii) ***The Almon DCC family:*** *in this case the intrinsic parameter subspace* $\mathcal{P}$ *is* $\mathbb{R}^3$ *and* $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (\mathrm{alm}_n(\boldsymbol{\theta}_1)(\mathrm{alm}_n(\boldsymbol{\theta}_1))^\top, \mathrm{alm}_n(\boldsymbol{\theta}_2)(\mathrm{alm}_n(\boldsymbol{\theta}_2))^\top)$, *with* $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}^3$, $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. *Moreover,*

$$\begin{array}{rccc} T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}: & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R}^3 \times \mathbb{R}^3 \\ & (\Delta_1, \Delta_2) & \longmapsto & 2\,(K^\top_{\boldsymbol{\theta}_1}\Delta_1 \mathrm{alm}_n(\boldsymbol{\theta}_1), K^\top_{\boldsymbol{\theta}_2}\Delta_2 \mathrm{alm}_n(\boldsymbol{\theta}_2)), \end{array} \tag{D.8}$$

where $K_{\boldsymbol{\theta}_i} = (\ \mathbf{i}_n \mid \mathbf{k}^1_n \odot \mathrm{alm}_n(\bar{\boldsymbol{\theta}}_i) \mid \mathbf{k}^2_n \odot \mathrm{alm}_n(\bar{\boldsymbol{\theta}}_i)) \in \mathbb{M}_{n,3}$, $\bar{\boldsymbol{\theta}}_i := (0, (\boldsymbol{\theta}_i)_2, (\boldsymbol{\theta}_i)_3)^\top$, $i \in \{1, 2\}$, *the symbol* $\mid$ *denotes vertical concatenation, and* $\mathbf{k}^1_n := (1, 2, \ldots, n)^\top$, $\mathbf{k}^2_n := \left(1, 2^2, \ldots, n^2\right)^\top \in \mathbb{R}^n$.

(iv) ***The scalar DCC family:*** *the intrinsic parameter subspace is* $\mathbb{R}$ *and* $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (a\mathbf{i}_n\mathbf{i}_n^\top, b\mathbf{i}_n\mathbf{i}_n^\top)$, *with* $a, b \in \mathbb{R}$, $\boldsymbol{\theta} = (a, b)$. *Moreover,*

$$\begin{array}{rccc} T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}: & \mathbb{S}_n \times \mathbb{S}_n & \longrightarrow & \mathbb{R} \times \mathbb{R} \\ & (\Delta_1, \Delta_2) & \longmapsto & \left(\langle \Delta_1, \mathbf{i}_n\mathbf{i}_n^\top \rangle, \langle \Delta_2, \mathbf{i}_n\mathbf{i}_n^\top \rangle\right). \end{array} \tag{D.9}$$

Remark D.2 Proposition D.1 can be easily extended to the non-targeted DCC model (2.5). In that case we take into account the presence of the additional intrinsic parameter $\boldsymbol{\theta} \in \mathbb{R}^N$, $N := \frac{1}{2}n\,(n+1)$, and hence we define $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \in \mathcal{P} \times \mathcal{P} \times \mathbb{R}^N$ and $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (A(\boldsymbol{\theta}_1), B(\boldsymbol{\theta}_2), C(\boldsymbol{\theta}_3)) \in \mathbb{S}_n \times \mathbb{S}_n \times \mathbb{S}_n$. The relations (D.1)-(D.2) remain valid, the maps corresponding to (D.3), (D.6)-(D.9) are obtained in a straightforward way by extending both their domain and image spaces to account for the added parameter $\boldsymbol{\theta}_3 \in \mathbb{R}^N$ applying the map (D.6). For the latter map the domain and image space have to be reduced from a product to one multiplier beforehand. The recursions(D.4)-(D.5) in the non-targeted case have to be extended in order to define a new differential operator $T^*_{\boldsymbol{\Theta}}Q_t : \mathbb{S}_n \times \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n \times \mathbb{S}_n$ that is determined by the recursions:

$$T^*_A Q_t \cdot \Delta = \Delta \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}^\top_{t-1}\right) + T^*_A Q_{t-1}\left[\Delta \odot B\right], \tag{D.10}$$

$$T^*_B Q_t \cdot \Delta = \Delta \odot Q_{t-1} + T^*_B Q_{t-1}\left[\Delta \odot B\right], \tag{D.11}$$

$$T^*_C Q_t \cdot \Delta = \Delta + T^*_C Q_{t-1}\left[\Delta \odot B\right]. \tag{D.12}$$

These recursions are initialized by setting $T^*_A Q_0 = 0$, $T^*_B Q_0 = 0$, and $T^*_C Q_0 = 0$.

Proof of Proposition D.1 We first compute the differential of the log-likelihood function in (3.2). Indeed, for any $\boldsymbol{\delta\theta} \in T_{\boldsymbol{\theta}}\mathcal{P} \times T_{\boldsymbol{\theta}}\mathcal{P}$:

$$\begin{aligned} d_{\boldsymbol{\theta}} l_t \cdot \boldsymbol{\delta\theta} &= d_{H_t} l_t\left(H_t\left(\boldsymbol{\Theta}\left(\boldsymbol{\theta}\right)\right)\right) \cdot T_{\boldsymbol{\Theta}} H_t\left(\boldsymbol{\Theta}\left(\boldsymbol{\theta}\right)\right) \cdot T_{\boldsymbol{\theta}}\boldsymbol{\Theta}\left(\boldsymbol{\theta}\right) \cdot \boldsymbol{\delta\theta} \\ &= \langle \nabla_{H_t} l_t, T_{\boldsymbol{\Theta}} H_t \cdot T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot \boldsymbol{\delta\theta} \rangle = \langle T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot T^*_{\boldsymbol{\Theta}} H_t \cdot \nabla_{H_t} l_t, \boldsymbol{\delta\theta} \rangle, \end{aligned}$$

where we used the chain rule on the function $\boldsymbol{\Theta} = \boldsymbol{\Theta}(\boldsymbol{\theta})$. This relation shows that

$$\nabla_{\boldsymbol{\theta}} l_t\left(\boldsymbol{\theta}; \mathbf{r}_t\right) = T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot T^*_{\boldsymbol{\Theta}} H_t \cdot \nabla_{H_t} l_t\left(\boldsymbol{\theta}; \mathbf{r}_t\right),$$

which yields (D.1). The proof of (D.2) is standard and can be found, for instance, in Apendix 7.6 of [CO14]. In order to prove (D.3) we first consider the tangent map $T_{\boldsymbol{\Theta}}H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ and derive the expression that determines it for every component Θ of $\boldsymbol{\Theta}$. First,

$$\begin{aligned} T_\Theta H_t \cdot \delta\Theta &= D_t(T_\Theta(Q_t^{*-1}) \cdot \delta\Theta)Q_tQ_t^{*-1}D_t + D_tQ_t^{*-1}(T_\Theta Q_t \cdot \delta\Theta)Q_t^{*-1}D_t \\ &\quad + D_tQ_t^{*-1}Q_t(T_\Theta(Q_t^{*-1}) \cdot \delta\Theta)D_t, \quad \delta\Theta \in \mathbb{S}_n. \end{aligned} \tag{D.13}$$

Now, since

$$T_\Theta Q_t^* \cdot \delta\Theta = \frac{1}{2}Q_t^{*-1}\mathrm{Diag}\left(T_\Theta Q_t \cdot \delta\Theta\right),$$

the following relation holds true

$$T_\Theta(Q_t^{*-1}) \cdot \delta\Theta = -\frac{1}{2}Q_t^{*-2}\mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)Q_t^{*-1}. \tag{D.14}$$

When we substitute the expression (D.14) into the relation (D.13) we obtain an explicit expression for the map $T_{\boldsymbol{\Theta}}H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$. We now compute its adjoint $T^*_{\boldsymbol{\Theta}}H_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ by dualizing (D.13) with respect to the Frobenius inner product. Indeed, let $\delta\Theta, \Delta \in \mathbb{S}_n$ arbitrary, then:

$$\begin{aligned} \langle T^*_\Theta H_t \cdot \Delta, \delta\Theta\rangle &= \langle \Delta, T_\Theta H_t \cdot \delta\Theta\rangle = \langle \Delta, D_t(T_\Theta(Q_t^{*-1}) \cdot \delta\Theta)Q_tQ_t^{*-1}D_t\rangle + \langle \Delta, D_tQ_t^{*-1}(T_\Theta Q_t \cdot \delta\Theta)Q_t^{*-1}D_t\rangle \\ &\quad + \langle \Delta, D_tQ_t^{*-1}Q_t\left(T_\Theta\left(Q_t^{*-1}\right) \cdot \delta\Theta\right)D_t\rangle = \langle T^*_\Theta\left(Q_t^{*-1}\right)\left(D_t\Delta D_tQ_t^{*-1}Q_t\right), \delta\Theta\rangle \\ &\quad + \langle T^*_\Theta Q_t\left(Q_t^{*-1}D_t\Delta D_tQ_t^{*-1}\right), \delta\Theta\rangle + \langle T^*_\Theta Q_t^{*-1}\left(Q_tQ_t^{*-1}D_t\Delta D_t\right), \delta\Theta\rangle, \end{aligned}$$

and hence

$$T^*_\Theta H_t \cdot \Delta = T^*_\Theta Q_t\left(Q_t^{*-1}D_t\Delta D_tQ_t^{*-1}\right) + T^*_\Theta\left(Q_t^{*-1}\right)\left[D_t\Delta D_tQ_t^{*-1}Q_t + Q_tQ_t^{*-1}D_t\Delta D_t\right]. \tag{D.15}$$

In order to explicitly write down the second summand of this relation, we dualize (D.14). For any $\delta\Theta, \Delta \in \mathbb{S}_n$ we obtain:

$$\begin{aligned} \langle T^*_\Theta(Q_t^{*-1})\Delta, \delta\Theta\rangle &= -\frac{1}{2}\mathrm{tr}(\Delta^\top Q_t^{*-2}\mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)Q_t^{*-1}) = -\frac{1}{2}\mathrm{tr}(Q_t^{*-1}\Delta^\top Q_t^{*-2}\mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)) \\ &= -\frac{1}{2}\langle Q_t^{*-2}\Delta Q_t^{*-1}, \mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)\rangle = -\frac{1}{2}\langle \mathrm{Diag}(Q_t^{*-2}\Delta Q_t^{*-1}), \mathrm{Diag}(T_\Theta Q_t \cdot \delta\Theta)\rangle \\ &= -\frac{1}{2}\langle T^*_\Theta Q_t\mathrm{Diag}(Q_t^{*-2}\Delta Q_t^{*-1}), \delta\Theta\rangle, \end{aligned}$$

which immediately yields

$$T^*_\Theta\left(Q_t^{*-1}\right) \cdot \Delta = -\frac{1}{2}T^*_\Theta Q_t\mathrm{Diag}\left(Q_t^{*-2}\Delta Q_t^{*-1}\right). \tag{D.16}$$

Substituting (D.16) into (D.15) we obtain:

$$T^*_\Theta H_t \cdot \Delta = T^*_\Theta Q_t\left[Q_t^{*-1}D_t\Delta D_tQ_t^{*-1} - \frac{1}{2}\mathrm{Diag}\left(Q_t^{*-2}\left(D_t\Delta D_tQ_t^{*-1}Q_t + Q_tQ_t^{*-1}D_t\Delta D_t\right)Q_t^{*-1}\right)\right],$$

which proves (D.3) in the statement of the Proposition.

In order to prove the relations (D.4) and (D.5) we start by differentiating (2.8). For arbitrary $\delta A, \delta B \in \mathbb{S}_n$ we obtain:

$$\begin{aligned} T_AQ_t \cdot \delta A &= -\delta A \odot S + \delta A \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top\right) + B \odot (T_AQ_{t-1} \cdot \delta A) \\ &= \delta A \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right) + B \odot (T_AQ_{t-1} \cdot \delta A), \end{aligned} \tag{D.17}$$

$$\begin{aligned} T_BQ_t \cdot \delta B &= -\delta B \odot S + \delta B \odot Q_{t-1} + B \odot (T_BQ_{t-1} \cdot \delta B) \\ &= \delta B \odot (Q_{t-1} - S) + B \odot (T_BQ_{t-1} \cdot \delta B). \end{aligned} \tag{D.18}$$

We now compute the corresponding adjoints with respect to the Frobenius inner product. Indeed, for any $\delta A, \Delta \in \mathbb{S}_n$ we have

$$\begin{aligned}\langle T_A^* Q_t \cdot \Delta, \delta A\rangle &= \operatorname{tr}\left[\Delta^\top \left(\delta A \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right)\right)\right] + \operatorname{tr}\left[\Delta^\top \left(B \odot (T_A Q_{t-1} \cdot \delta A)\right)\right] \\ &= \operatorname{tr}\left[\delta A^\top \left(\Delta \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right)\right)\right] + \operatorname{tr}\left[\left(T_A Q_{t-1} \cdot \delta A\right)^\top (\Delta \odot B)\right] \\ &= \left\langle \Delta \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right), \delta A\right\rangle + \left\langle T_A^* Q_{t-1} (\Delta \odot B), \delta A\right\rangle,\end{aligned}$$

where we used the Hadamard product trace property (A.4). This expression yields

$$T_A^* Q_t \cdot \Delta = \Delta \odot \left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right) + T_A^* Q_{t-1} (\Delta \odot B)$$

which proves (D.4). We analogously show (D.5) by dualizing (D.18). Let $\delta B, \Delta \in \mathbb{S}_n$ be arbitrary; the following relation holds true:

$$\langle T_B^* Q_t \cdot \Delta, \delta B\rangle = \operatorname{tr}\left[\Delta^\top (\delta B \odot (Q_{t-1} - S))\right] = \operatorname{tr}\langle \Delta \odot (Q_{t-1} - S), \delta B\rangle + \langle T_B^* Q_{t-1} (\Delta \odot B), \delta B\rangle,$$

which is equivalent to $T_B^* Q_t \cdot \Delta = \Delta \odot (Q_{t-1} - S) + T_B^* Q_{t-1} (\Delta \odot B)$.

We conclude by proving the relations (D.6)-(D.9) that are obtained out of the dualization of the tangent map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta}$ of the relation $\boldsymbol{\Theta}(\boldsymbol{\theta})$ that provides the connection between the variables $\boldsymbol{\Theta}$ and the intrinsic parameter space $\boldsymbol{\theta}$ for each of the families considered in Section 2.

(i) ***The Hadamard family:*** let $\boldsymbol{\delta a}, \boldsymbol{\delta b} \in \mathbb{R}^N$, then the map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is given by

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b}) = (\operatorname{math}(\boldsymbol{\delta a}), \operatorname{math}(\boldsymbol{\delta b})).$$

Dualizing this relation, we obtain for any $\Delta_1, \Delta_2 \in \mathbb{S}_n$:

$$\langle T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta}(\Delta_1, \Delta_2), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = \langle (\Delta_1, \Delta_2), (\operatorname{math}(\boldsymbol{\delta a}), \operatorname{math}(\boldsymbol{\delta b}))\rangle = \langle (\operatorname{math}^*(\Delta_1), \operatorname{math}^*(\Delta_2), (\boldsymbol{\delta a}, \boldsymbol{\delta b}))\rangle,$$

and hence

$$T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta}(\Delta_1, \Delta_2) = (\operatorname{math}^*(\Delta_1), \operatorname{math}^*(\Delta_2)),$$

which yields (D.6).

(ii) ***The rank deficient family:*** let $\boldsymbol{\delta a}, \boldsymbol{\delta b} \in \mathbb{R}^{N^*}$, the tangent map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is given by

$$\begin{aligned}T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b}) = \big(&\operatorname{mat}_r(\boldsymbol{\delta a})(\operatorname{mat}_r(\boldsymbol{a}))^\top + \operatorname{mat}_r(\boldsymbol{a})(\operatorname{mat}_r(\boldsymbol{\delta a}))^\top, \\ &\operatorname{mat}_r(\boldsymbol{\delta b})(\operatorname{mat}_r(\boldsymbol{b}))^\top + \operatorname{mat}_r(\boldsymbol{b})(\operatorname{mat}_r(\boldsymbol{\delta b}))^\top\big).\end{aligned}$$

We now dualize this relation in order to prove (D.7). For $\Delta_1, \Delta_2 \in \mathbb{S}_n$ arbitrary we write

$$\begin{aligned}\langle T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta}(\Delta_1, \Delta_2), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle &= \langle (\Delta_1, \Delta_2), T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle \\ &= \langle (\Delta_1, \Delta_2), (\operatorname{mat}_r(\boldsymbol{\delta a})(\operatorname{mat}_r(\boldsymbol{a}))^\top, \operatorname{mat}_r(\boldsymbol{\delta b})(\operatorname{mat}_r(\boldsymbol{b}))^\top)\rangle \\ &\quad + \langle (\Delta_1, \Delta_2), (\operatorname{mat}_r(\boldsymbol{a})(\operatorname{mat}_r(\boldsymbol{\delta a}))^\top, \operatorname{mat}_r(\boldsymbol{b})(\operatorname{mat}_r(\boldsymbol{\delta b}))^\top)\rangle \\ &= 2\langle (\Delta_1 \operatorname{mat}_r(\boldsymbol{a}), \Delta_2 \operatorname{mat}_r(\boldsymbol{b})), (\operatorname{mat}_r(\boldsymbol{\delta a}), \operatorname{mat}_r(\boldsymbol{\delta b}))\rangle \\ &= 2\langle (\operatorname{mat}_r^*(\Delta_1 \operatorname{mat}_r(\boldsymbol{a})), \operatorname{mat}_r^*(\Delta_2 \operatorname{mat}_r(\boldsymbol{b})), (\boldsymbol{\delta a}, \boldsymbol{\delta b}))\rangle. \qquad \text{(D.19)}\end{aligned}$$

Recall that by Proposition A.13 the operator mat_r^* equals vec_r and hence the relation (D.19) yields

$$T_{\boldsymbol{\theta}}^* \boldsymbol{\Theta}(\Delta_1, \Delta_2) = 2(\operatorname{vec}_r(\Delta_1 \operatorname{mat}_r(\boldsymbol{a})), \operatorname{vec}_r(\Delta_2 \operatorname{mat}_r(\boldsymbol{b}))),$$

and proves (D.7), as required.

(iii) ***The Almon family:*** let $\boldsymbol{\delta a}, \boldsymbol{\delta b} \in \mathbb{R}^3$, then the map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is determined by

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b}) = \Big((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})(\mathrm{alm}_n(\boldsymbol{a}))^\top + \mathrm{alm}_n(\boldsymbol{a})(K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top, \\ (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})(\mathrm{alm}_n(\boldsymbol{b}))^\top + \mathrm{alm}_n(\boldsymbol{b})(K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top \Big), \tag{D.20}$$

where we used the expression (A.16) for the tangent map $T_{\boldsymbol{\theta}}\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$. In order to prove (D.8) we dualize the relation (D.20). For $\Delta_1, \Delta_2 \in \mathbb{S}_n$ arbitrary we compute

$$\begin{aligned} \langle T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}(\Delta_1, \Delta_2), (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \rangle &= \langle (\Delta_1, \Delta_2), T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \rangle \\ &= \langle (\Delta_1, \Delta_2), ((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})(\mathrm{alm}_n(\boldsymbol{a}))^\top, (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})(\mathrm{alm}_n(\boldsymbol{b}))^\top) \rangle \\ &\quad + \langle (\Delta_1, \Delta_2), (\mathrm{alm}_n(\boldsymbol{a})(K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top, \mathrm{alm}_n(\boldsymbol{b})(K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top) \rangle \\ &= 2\langle (\Delta_1 \mathrm{alm}_n(\boldsymbol{a}), \Delta_2 \mathrm{alm}_n(\boldsymbol{b})), (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}, K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \rangle \\ &= 2\langle \big(K_{\boldsymbol{a}}^\top \cdot \Delta_1 \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top \cdot \Delta_2 \cdot \mathrm{alm}_n(\boldsymbol{b})\big), (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \rangle, \end{aligned}$$

which yields (D.8), as required.

(iv) ***The scalar family:*** let $\delta a, \delta b \in \mathbb{R}$. The map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ is given by

$$T_{\boldsymbol{\theta}}\boldsymbol{\Theta} \cdot (\delta a, \delta b) = \big(\delta a \mathbf{i}_n \mathbf{i}_n^\top, \delta b \mathbf{i}_n \mathbf{i}_n^\top\big).$$

Hence, for $\Delta_1, \Delta_2 \in \mathbb{S}_n$ arbitrary we have

$$\begin{aligned} \langle T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}(\Delta_1, \Delta_2), (\delta a, \delta b) \rangle &= \langle (\Delta_1, \Delta_2), T_{\boldsymbol{\theta}}\boldsymbol{\Theta}(\delta a, \delta b) \rangle \\ &= \langle (\Delta_1, \Delta_2), \big(\delta a \mathbf{i}_n \mathbf{i}_n^\top, \delta b \mathbf{i}_n \mathbf{i}_n^\top\big) \rangle = \langle \big(\langle \Delta_1, \mathbf{i}_n \mathbf{i}_n^\top \rangle, \langle \Delta_2, \mathbf{i}_n \mathbf{i}_n^\top \rangle\big), (\delta a, \delta b) \rangle, \end{aligned}$$

and consequently,

$$T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta}(\Delta_1, \Delta_2) = \big(\langle \Delta_1, \mathbf{i}_n \mathbf{i}_n^\top \rangle, \langle \Delta_2, \mathbf{i}_n \mathbf{i}_n^\top \rangle\big),$$

which proves (D.9), as required.

D.2 Matrix expressions of the recursions in Proposition D.1

In order to algorithmically implement Proposition D.1, the operator recursions (D.3)-(D.8) have to be expressed in terms of matrix recursions that are provided in the following result.

Proposition D.3 Consider $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (A(\boldsymbol{\theta}_1), B(\boldsymbol{\theta}_2))$, and the differential operators $T^*_{\boldsymbol{\Theta}}H_t$, $T^*_{\boldsymbol{\Theta}}Q_t : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$, which for every component Θ of $\boldsymbol{\Theta}$ are defined by the relations (D.3)-(D.5). Let $\mathcal{A}^t_Q, \mathcal{B}^t_Q, \mathcal{V}^t_\Theta : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ with $N := \frac{1}{2}n(n+1)$ be the linear maps, defined as

$$\mathcal{A}^t_Q := \mathrm{vech} \circ T^*_A Q_t \circ \mathrm{math}, \tag{D.21}$$

$$\mathcal{B}^t_Q := \mathrm{vech} \circ T^*_B Q_t \circ \mathrm{math}, \tag{D.22}$$

$$\mathcal{V}^t_\Theta := \mathrm{vech} \circ T^*_\Theta H_t \circ \mathrm{math}, \tag{D.23}$$

and let $A_t, B_t, V_t \in \mathbb{S}_N$ be the matrices associated to the operators (D.21), (D.22), and (D.23), respectively. Then, the matrices $\{A_t\}_{t\in\{1,\ldots,T\}}$, $\{B_t\}_{t\in\{1,\ldots,T\}}$, and $\{V_t\}_{t\in\{1,\ldots,T\}}$ are determined by the recursions

$$A_t = \mathrm{diag}\big(\mathrm{vech}\big(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\big)\big) + A_{t-1}\mathrm{diag}(\mathrm{vech}(B)), \tag{D.24}$$

$$B_t = \mathrm{diag}(\mathrm{vech}(Q_{t-1} - S)) + B_{t-1}\mathrm{diag}(\mathrm{vech}(B)), \tag{D.25}$$

$$\begin{aligned} V_t = \Theta_t \Big\{ & L_n\big((Q_t^{*-1}D_t) \otimes (Q_t^{*-1}D_t)\big) - \frac{1}{2}P_n^d L_n((Q_t^{*-1}Q_tQ_t^{*-1}D_t) \otimes (Q_t^{*-2}D_t) \\ & + (Q_t^{*-1}D_t) \otimes (Q_t^{*-2}Q_tQ_t^{*-1}D_t)) \Big\} \cdot D_n \ \text{ with } \ \Theta_t = \{A_t, B_t\}, \end{aligned} \tag{D.26}$$

and the initial values $A_0 = B_0 = V_0 = 0$, where L_n, D_n, and P_n^d are the elimination, duplication, and diagonalization matrices in dimension n, respectively, defined in (A.7)-(A.9).
Finally, consider the map $\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathcal{P} \times \mathcal{P}$ defined by

$$\mathcal{G}_{\boldsymbol{\theta}} := T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} \circ \text{math}, \tag{D.27}$$

where $T^*_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathcal{P} \times \mathcal{P}$ is the adjoint of the map $T_{\boldsymbol{\theta}}\boldsymbol{\Theta} : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{S}_n \times \mathbb{S}_n$ and $\mathcal{P} \times \mathcal{P}$ the intrinsic $\boldsymbol{\theta}$ parameter space that depends on the mapping $\boldsymbol{\Theta}(\boldsymbol{\theta})$ associated to each of the model specifications considered in Proposition D.1. Then:

(i) ***The Hadamard family:***

$$\begin{aligned}
&\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{R}^N \times \mathbb{R}^N \\
&\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{v}_1, \mathbf{v}_2) = 2\left(\mathbf{v}_1 - \frac{1}{2}\text{vech}(\text{Diag}(\text{math}(\mathbf{v}_1))), \mathbf{v}_2 - \frac{1}{2}\text{vech}(\text{Diag}(\text{math}(\mathbf{v}_2)))\right).
\end{aligned} \tag{D.28}$$

(ii) ***The rank deficient family:***

$$\begin{aligned}
&\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \\
&\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{v}_1, \mathbf{v}_2) = 2\,(\text{vec}_r(\text{math}(\mathbf{v}_1)\text{mat}_r(\boldsymbol{\theta}_1)), \text{vec}_r(\text{math}(\mathbf{v}_2)\text{mat}_r(\boldsymbol{\theta}_2))),
\end{aligned} \tag{D.29}$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$, and where $r \leq n \in \mathbb{N}$, $N^* := nr - \frac{1}{2}r(r-1)$.

(iii) ***The Almon family:***

$$\begin{aligned}
&\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{R}^3 \times \mathbb{R}^3 \\
&\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{v}_1, \mathbf{v}_2) = 2\,(K^\top_{\boldsymbol{\theta}_1}\text{math}(\mathbf{v}_1)\text{alm}_n(\boldsymbol{\theta}_1), K^\top_{\boldsymbol{\theta}_2}\text{math}(\mathbf{v}_2)\text{alm}_n(\boldsymbol{\theta}_2)),
\end{aligned} \tag{D.30}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^3 \times \mathbb{R}^3$, and $K_{\boldsymbol{\theta}_i}$ (for $i = 1$ and 2) has been defined in Proposition D.1.

(iv) ***The scalar family:***

$$\begin{aligned}
&\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{R} \times \mathbb{R} \\
&\mathcal{G}_{\boldsymbol{\theta}}(\mathbf{v}_1, \mathbf{v}_2) = \left(\langle \text{math}(\mathbf{v}_1), \mathbf{i}_n\mathbf{i}_n^\top\rangle, \langle \text{math}(\mathbf{v}_2), \mathbf{i}_n\mathbf{i}_n^\top\rangle\right).
\end{aligned} \tag{D.31}$$

Proposition D.3 provides a complete computational recipe for the calculation of the gradient of the log-likelihood function associated to each of the DCC model parameterizations under study. We note that this result could be easily extended to any other model prescription by simply writing down the associated operator $\mathcal{G}_{\boldsymbol{\theta}} : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathcal{P} \times \mathcal{P}$ in (D.27).

Remark D.4 Proposition D.3 can be extended to the non-targeted DCC model (2.5). In that case $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ and $\boldsymbol{\Theta}(\boldsymbol{\theta}) := (A(\boldsymbol{\theta}_1), B(\boldsymbol{\theta}_2), C(\boldsymbol{\theta}_3))$. The differential operators $T^*_{\boldsymbol{\Theta}}H_t$, $T^*_{\boldsymbol{\Theta}}Q_t : \mathbb{S}_n \times \mathbb{S}_n \times \mathbb{S}_n \longrightarrow \mathbb{S}_n \times \mathbb{S}_n \times \mathbb{S}_n$, for every component Θ of $\boldsymbol{\Theta}$ are defined by the relations (D.3) and (D.10)-(D.12). The maps $\mathcal{A}^t_Q, \mathcal{B}^t_Q, \mathcal{V}^t_\Theta : \mathbb{R}^N \longrightarrow \mathbb{R}^N$, with $N := \frac{1}{2}n(n+1)$, are defined as in (D.21)-(D.23) and, additionally, $\mathcal{C}^t_Q : \mathbb{R}^N \to \mathbb{R}^N$ is given by

$$\mathcal{C}^t_Q := \text{vech} \circ T^*_C Q_t \circ \text{math}. \tag{D.32}$$

The matrices $A_t, B_t, C_t, V_t \in \mathbb{S}_N$ associated to the operators (D.21), (D.22), (D.32), and (D.23), respectively, are determined by the recursions:

$$\begin{aligned}
A_t &= \operatorname{diag}\left(\operatorname{vech}\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top\right)\right) + A_{t-1}\operatorname{diag}\left(\operatorname{vech}\left(B\right)\right), && \text{(D.33)}\\
B_t &= \operatorname{diag}\left(\operatorname{vech}\left(Q_{t-1}\right)\right) + B_{t-1}\operatorname{diag}\left(\operatorname{vech}\left(B\right)\right), && \text{(D.34)}\\
C_t &= \mathbb{I}_N + C_{t-1}\operatorname{diag}\left(\operatorname{vech}\left(B\right)\right), && \text{(D.35)}\\
V_t &= \Theta_t\Big\{L_n\left(\left(Q_t^{*-1}D_t\right)\otimes\left(Q_t^{*-1}D_t\right)\right) - \frac{1}{2}P_n^d L_n((Q_t^{*-1}Q_tQ_t^{*-1}D_t)\otimes(Q_t^{*-2}D_t)\\
&\quad +(Q_t^{*-1}D_t)\otimes(Q_t^{*-2}Q_tQ_t^{*-1}D_t))\Big\}\cdot D_n \text{ with } \Theta_t = \{A_t, B_t, C_t\}, && \text{(D.36)}
\end{aligned}$$

using the initial values $A_0 = B_0 = C_0 = V_0 = 0$; L_n, D_n, and P_n^d are the elimination, duplication, and diagonalization matrices in dimension n, respectively, defined in (A.7)-(A.9). The generalizations of the maps (D.27)-(D.31) to the non-targeted case are obtained in a straightforward way by extending both the domain and image spaces in order to account for $\boldsymbol{\theta}_3 \in \mathbb{R}^N$, $N := \frac{1}{2}n\,(n+1)$. For that these both corresponding to $\boldsymbol{\theta}_3$ domain and image subspaces are taken from the definition (D.28) having reduced it first to a one-component map.

Proof of Proposition D.3

We start by proving (D.24). Let $\Delta := \operatorname{math}(\mathbf{v})$, with $\mathbf{v} \in \mathbb{R}^N$ arbitrary. Then by (D.4), recalling that the operator math is the inverse of the operator vech, and using the expression (A.3) for the Hadamard product of two vectors, we obtain

$$\begin{aligned}
A_t\cdot\mathbf{v} &= \operatorname{vech}\left(\operatorname{math}(\mathbf{v})\odot\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right)\right) + \operatorname{vech}\, T_A^* Q_{t-1}\,\operatorname{math}\,\operatorname{vech}\left(\operatorname{math}(\mathbf{v})\odot B\right)\\
&= \operatorname{vech}\left(\operatorname{math}(\mathbf{v})\odot\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right)\right) + A_{t-1}\,\operatorname{vech}\left(\operatorname{math}(\mathbf{v})\odot B\right)\\
&= \mathbf{v}\odot\operatorname{vech}\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right) + A_{t-1}\left(\mathbf{v}\odot\operatorname{vech}\left(B\right)\right)\\
&= \operatorname{diag}\left(\operatorname{vech}\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^\top - S\right)\right)\cdot\mathbf{v} + A_{t-1}\operatorname{diag}\left(\operatorname{vech}\left(B\right)\right)\cdot\mathbf{v},
\end{aligned}$$

which yields (D.24) as required. Analogously, by (D.5):

$$\begin{aligned}
B_t\cdot\mathbf{v} &= \operatorname{vech}\left(\operatorname{math}(\mathbf{v})\odot\left(Q_{t-1} - S\right)\right) + \operatorname{vech}\, T_B^* Q_{t-1}\,\operatorname{math}\,\operatorname{vech}\left(\operatorname{math}(\mathbf{v})\odot B\right)\\
&= \mathbf{v}\odot\operatorname{vech}\left(Q_{t-1} - S\right) + B_{t-1}\left(\mathbf{v}\odot\operatorname{vech}\left(B\right)\right) = \operatorname{diag}\left(\operatorname{vech}\left(Q_{t-1} - S\right)\right)\cdot\mathbf{v} + A_{t-1}\operatorname{diag}\left(\operatorname{vech}\left(B\right)\right)\cdot\mathbf{v},
\end{aligned}$$

and hence (D.25) follows.

We now prove (D.26) out of (D.3) and apply the relations (A.7)-(A.10) between the vech, math, vec, and Diag:

$$\begin{aligned}
V_t\cdot\mathbf{v} &= \operatorname{vech}\, T_\Theta^* Q_t\,\operatorname{math}\,\operatorname{vech}\left(Q_t^{*-1}D_t\,\operatorname{math}(\mathbf{v})\,D_tQ_t^{*-1}\right)\\
&\quad -\frac{1}{2}\operatorname{vech}\, T_\Theta^* Q_t\,\operatorname{math}\,\operatorname{vech}\,\operatorname{Diag}(Q_t^{*-2}D_t\,\operatorname{math}(\mathbf{v})\,D_tQ_t^{*-1}Q_tQ_t^{*-1} + Q_t^{*-2}Q_tQ_t^{*-1}D_t\operatorname{math}(\mathbf{v})\,D_tQ_t^{*-1})\\
&= \Theta_tL_n\operatorname{vec}\left(Q_t^{*-1}D_t\,\operatorname{math}(\mathbf{v})\,D_tQ_t^{*-1}\right) - \frac{1}{2}\Theta_tL_n\operatorname{vec}\,\operatorname{Diag}(Q_t^{*-2}D_t\,\operatorname{math}(\mathbf{v})\,D_tQ_t^{*-1}Q_tQ_t^{*-1}\\
&\quad +Q_t^{*-2}Q_tQ_t^{*-1}D_t\operatorname{math}(\mathbf{v})\,D_tQ_t^{*-1}) = \Theta_tL_n\left(\left(Q_t^{*-1}D_t\right)\otimes\left(Q_t^{*-1}D_t\right)\right)D_n\cdot\mathbf{v}\\
&\quad -\frac{1}{2}\Theta_tL_n\operatorname{vec}\,\operatorname{math}\,P_n^dL_n\operatorname{vec}(Q_t^{*-2}D_t\,\operatorname{math}(\mathbf{v})\,D_tQ_t^{*-1}Q_tQ_t^{*-1} + Q_t^{*-2}Q_tQ_t^{*-1}D_t\operatorname{math}(\mathbf{v})\,D_tQ_t^{*-1})\\
&= \Theta_tL_n\left(\left(Q_t^{*-1}D_t\right)\otimes\left(Q_t^{*-1}D_t\right)\right)D_n\cdot\mathbf{v} - \frac{1}{2}\Theta_tP_n^dL_n((Q_t^{*-1}Q_tQ_t^{*-1}D_t)\otimes\left(Q_t^{*-2}D_t\right)\\
&\quad +\left(Q_t^{*-1}D_t\right)\otimes\left(Q_t^{*-2}Q_tQ_t^{*-1}D_t\right))D_n\cdot\mathbf{v},
\end{aligned}$$

with $\Theta_t = \{A_t, B_t\}$, which proves (D.26) as required. Finally, the relations (D.28)-(D.31) are obtained from (D.6)-(D.9) by straightforward substitution.

E Implementation of the constrained optimization method in the estimation of DCC models

The local optimization problem in (3.7), i.e. $\boldsymbol{\theta}^{(k+1)} = \arg\min_{\boldsymbol{\theta}\in\mathcal{P}\times\mathcal{P}} \tilde{f}^{(k)}(\boldsymbol{\theta})$, is solved by finding the value $\boldsymbol{\theta}_0$ for which $\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta}_0) = 0$. The general expression of the gradient $\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ of the local model is given by

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta}) &= \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) + H^{(k)}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{s_1} L_1^j \nabla_{\boldsymbol{\theta}} D_{M_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) \\ &\quad + \sum_{j=1}^{s_2} L_2^j \sum_{i=1}^{q_j} \nabla_{\boldsymbol{\theta}} D^i_{N_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{s_3} L_3^j \sum_{i=1}^{m_j} \nabla_{\boldsymbol{\theta}} D^i_{L_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}). \end{aligned} \tag{E.1}$$

The gradients of the divergences in (E.1) are given by

$$\nabla_\theta D_{M_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = -T^*_\theta M_j(M_j(\boldsymbol{\theta})^{-1} - M_j(\boldsymbol{\theta}^{(k)})^{-1}), \tag{E.2}$$

$$\nabla_\theta D^i_{N_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = -\left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right](T^*_\theta N_j \cdot \mathbf{e}^{(i)}_{q_j}),\ \ i\in\{1,\dots,q_j\}, \tag{E.3}$$

$$\nabla_\theta D^i_{L_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = \left[\frac{1}{(L_j(\boldsymbol{\theta}))_i} - \frac{1}{(L_j(\boldsymbol{\theta}^{(k)}))_i}\right](C^{(j)T}_\theta \cdot \mathbf{e}^{(i)}_{m_j}),\ \ i\in\{1,\dots,m_j\}, \tag{E.4}$$

where $\theta\in\mathcal{P}$ is a generic component of the intrinsic parameter $\boldsymbol{\theta}\in\mathcal{P}\times\mathcal{P}$, $C^{(j)}_\theta \in \mathbb{M}_{m_j,S}$, $j\in\{1,\dots,s_3\}$ is the matrix associated to the j-th linear constraint, and $\mathbf{e}^{(i)}_{q_j}\in\mathbb{R}^{q_j}$, $i\in\{1,\dots,q_j\}$ and $\mathbf{e}^{(i)}_{m_j}\in\mathbb{R}^{m_j}$, $i\in\{1,\dots,m_j\}$ are the canonical unit vectors introduced in Subsection A.1. Additionally, for every $j\in\{1,\dots,s_1\}$, $T^*_{\boldsymbol{\theta}}M_j : \mathbb{S}_q \longrightarrow \mathcal{P}\times\mathcal{P}$ is the adjoint of the tangent map $T_{\boldsymbol{\theta}}M_j : \mathcal{P}\times\mathcal{P}\longrightarrow\mathbb{S}_q$ of $M_j(\boldsymbol{\theta}) : \mathcal{P}\times\mathcal{P}\longrightarrow\mathbb{S}_q$ that determines the j-th positive semidefinite constraint. The symbol $T^*_\theta M_j : \mathbb{S}_q\longrightarrow\mathcal{P}$ in the relation (E.2) denotes the adjoint of the partial derivative $T_\theta M_j : \mathcal{P}\longrightarrow\mathbb{S}_q$ for some component θ of $\boldsymbol{\theta}$. Analogously, for every $j\in\{1,\dots,s_2\}$, the map $T^*_{\boldsymbol{\theta}}N_j : \mathbb{R}^{q_j}\times\mathbb{R}^{q_j}\longrightarrow\mathcal{P}\times\mathcal{P}$ is the adjoint of the tangent map $T_{\boldsymbol{\theta}}N_j : \mathcal{P}\times\mathcal{P}\longrightarrow\mathbb{R}^{q_j}\times\mathbb{R}^{q_j}$ of the function $N_j(\boldsymbol{\theta}) : \mathcal{P}\times\mathcal{P}\longrightarrow\mathbb{R}^{q_j}$ that determines the j-th nonlinear constraint. Again $T^*_\theta N_j$ denotes the adjoint of the corresponding θ partial derivative.

In order to numerically obtain the solution $\boldsymbol{\theta}_0$ of $\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta}_0)=0$ using the Newton-Raphson algorithm, we need the Jacobian of the gradient of the local model and in particular the tangent maps of all the gradients of the divergences in (E.2)-(E.4). Straightforward computations yield for every component $\theta\in\mathcal{P}$ of $\boldsymbol{\theta}$ and any $\delta\theta\in\mathcal{P}$:

$$\begin{aligned} T_\theta(\nabla_\theta D_{M_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}))\cdot\delta\theta &= T^*_\theta M_j(M_j(\boldsymbol{\theta})^{-1}(T_\theta M_j\cdot\delta\theta)M_j(\boldsymbol{\theta})^{-1})) \\ &\quad - (T_\theta(T^*_\theta M_j))\,(M_j(\boldsymbol{\theta})^{-1} - M_j(\boldsymbol{\theta}^{(k)})^{-1})\cdot\delta\theta,\ \ j\in\{1,\dots,s_1\}, \\ T_\theta(\nabla_\theta D^i_{N_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}))\cdot\delta\theta &= \frac{(T_\theta N_j(\boldsymbol{\theta}))_i\cdot\delta\theta}{(N_j(\boldsymbol{\theta}))_i^2}(T^*_\theta N_j\cdot\mathbf{e}^{(i)}_{q_j}) - \left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right] T_\theta(T^*_\theta N_j\cdot\mathbf{e}^{(i)}_{q_j})\cdot\delta\theta, \\ &\qquad\qquad i\in\{1,\dots,q_j\},\ \ j\in\{1,\dots,s_2\}, \\ T_\theta(\nabla_\theta D^i_{L_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})\cdot\delta\theta &= \frac{(C^{(j)}_\theta\cdot\delta\theta)_i}{(L_j(\boldsymbol{\theta}))_i^2}\,(C^{(j)T}_\theta\cdot\mathbf{e}^{(i)}_{m_j}),\quad i\in\{1,\dots,m_j\},\ \ j\in\{1,\dots,s_3\}. \end{aligned}$$

Using these identities it is easy to determine the Jacobian of the gradient of the local model $\tilde{f}^{(k)}(\boldsymbol{\theta})$.

Indeed, the expression for the block of the Jacobian that corresponds to the component $\theta \in \mathcal{P}$ is:

$$\begin{aligned} T_\theta(\nabla_\theta \tilde{f}^{(k)}(\boldsymbol{\theta}))\cdot\delta\theta \;\; &= \;\; H^{(k)}\cdot\delta\theta + \sum_{j=1}^{s_1} L_1^j \Big[T_\theta^* M_j (M_j(\boldsymbol{\theta})^{-1}(T_\theta M_j\cdot\delta\theta)M_j(\boldsymbol{\theta})^{-1})) \\ &-(T_\theta(T_\theta^* M_j))(M_j(\boldsymbol{\theta})^{-1} - M_j(\boldsymbol{\theta}^{(k)})^{-1})\cdot\delta\theta\Big] + \sum_{j=1}^{s_2} L_2^j \sum_{i=1}^{q_j} \Big[\frac{(T_\theta N_j(\boldsymbol{\theta}))_i\cdot\delta\theta}{(N_j(\boldsymbol{\theta}))_i^2}(T_\theta^* N_j\cdot \mathbf{e}_{q_j}^{(i)}) \\ &- \left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right] T_\theta(T_\theta^* N_j\cdot\mathbf{e}_{q_j}^{(i)})\cdot\delta\theta\Big] + \sum_{j=1}^{s_3} L_3^j \sum_{i=1}^{m_j} \frac{(C_\theta^{(j)}\cdot\delta\theta)_i}{(L_j(\boldsymbol{\theta}))_i^2}\;(C_\theta^{(j)\top}\cdot\mathbf{e}_{m_j}^{(i)}). \end{aligned} \tag{E.5}$$

In the following subsections we provide explicit expressions for the Bregman divergences, the local penalized model, its gradient and associated Jacobian, for the different specifications of the DCC model that are presented in Section 2. The detailed derivations of these results are contained in Section F.

E.1 Implementation for the Hadamard DCC model

Remember that in this particular case, the matrices $A, B \in \mathbb{S}_n$ are parametrized with two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^N$, $N = \frac{1}{2}n(n+1)$, by setting $A := \text{math}(\boldsymbol{a})$, $B := \text{math}(\boldsymbol{b})$. Let $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^N \times \mathbb{R}^N$; the intrinsic parameter subspace $\mathcal{P}$ is in this case $\mathbb{R}^N$ and has dimension $P = N$. The dynamics of the conditional correlation matrix for the Hadamard DCC model is given by (2.12) and $\boldsymbol{\theta}$ is subjected to positivity constraints (2.13)-(2.14).

Constraints and the local model The positivity constraints that need to be imposed on the Hadamard model specification can be formulated using the classification introduced in Subsection 3.2.2 as the following set of positive semidefiniteness/definiteness conditions:

$$M_1(\boldsymbol{a}) := \text{math}(\boldsymbol{a}) \succeq 0, \tag{E.6}$$
$$M_2(\boldsymbol{b}) := \text{math}(\boldsymbol{b}) \succeq 0, \tag{E.7}$$
$$M_3(\boldsymbol{\theta}) := \left(\mathbf{i}_n\mathbf{i}_n^\top - \text{math}(\boldsymbol{a}) - \text{math}(\boldsymbol{b})\right) \odot S \succ 0, \tag{E.8}$$

where $M_1(\boldsymbol{a}), M_2(\boldsymbol{b}), M_3(\boldsymbol{\theta}) \in \mathbb{S}_n$. The local model corresponding to (3.7) is given in this case by

$$\tilde{f}^{(k)}(\boldsymbol{\theta}) = f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{3} L_1^j D_{M_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}), \tag{E.9}$$

where $\mathbf{L}_1 := \left(L_1^1, L_1^2, L_1^3\right)^\top$ is the vector that contains the penalization strengths, $f(\boldsymbol{\theta}^{(k)})$ is minus the log-likelihood function (3.1), $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})$ is its gradient, that is $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \text{log}L(\boldsymbol{\theta}^{(k)}; \mathbf{r})$, which for the Hadamard model is determined by relations (D.1)-(D.6) in Proposition D.1, and $H^{(k)}$ is its Hessian computed at the point $\boldsymbol{\theta}^{(k)}$. In relation (E.9), the Bregman divergences associated to each of the constraints (E.6)-(E.8) are easily obtained from (3.5), which yields the following expressions:

$$D_{M_1}(\boldsymbol{a}, \boldsymbol{a}^{(k)}) = \text{tr}(M_1(\boldsymbol{a})\cdot M_1(\boldsymbol{a}^{(k)})^{-1}) - \log\det(M_1(\boldsymbol{a})\cdot M_1(\boldsymbol{a}^{(k)})^{-1}) - n, \tag{E.10}$$
$$D_{M_2}(\boldsymbol{b}, \boldsymbol{b}^{(k)}) = \text{tr}(M_2(\boldsymbol{b})\cdot M_2(\boldsymbol{b}^{(k)})^{-1}) - \log\det(M_2(\boldsymbol{b})\cdot M_2(\boldsymbol{b}^{(k)})^{-1}) - n, \tag{E.11}$$
$$D_{M_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \text{tr}(M_3(\boldsymbol{a}, \boldsymbol{b})\cdot M_3(\boldsymbol{a}^{(k)}, \boldsymbol{b}^{(k)})^{-1}) - \log\det(M_3(\boldsymbol{a}, \boldsymbol{b})\cdot M_3(\boldsymbol{a}^{(k)}, \boldsymbol{b}^{(k)})^{-1}) - n. \tag{E.12}$$

Gradient of the local model A lengthy but straightforward computation contained in the technical appendix F.1 yields the following expressions for the components of the gradient of the local model:

$$\begin{aligned}\nabla_{\boldsymbol{a}}\tilde{f}^{(k)}(\boldsymbol{\theta}) =& \nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{a}}^{(k)}(\boldsymbol{a}-\boldsymbol{a}^{(k)}) - L_1^1\ \mathrm{math}^*(\mathrm{math}(\boldsymbol{a})^{-1} - \mathrm{math}(\boldsymbol{a}^{(k)})^{-1}) \\ &+ L_1^3\ \mathrm{math}^*\left((Z^{-1} - (Z^{(k)})^{-1}) \odot S\right), \end{aligned} \tag{E.13}$$

$$\begin{aligned}\nabla_{\boldsymbol{b}}\tilde{f}^{(k)}(\boldsymbol{\theta}) =& \nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{b}}^{(k)}(\boldsymbol{b}-\boldsymbol{b}^{(k)}) - L_1^2\ \mathrm{math}^*(\mathrm{math}(\boldsymbol{b})^{-1} - \mathrm{math}(\boldsymbol{b}^{(k)})^{-1}) \\ &+ L_1^3\ \mathrm{math}^*\left((Z^{-1} - (Z^{(k)})^{-1}) \odot S\right), \end{aligned} \tag{E.14}$$

where we used the conventions $Z := \left(\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}) - \mathrm{math}(\boldsymbol{b})\right) \odot S$ and $Z^{(k)} := (\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{math}(\boldsymbol{a}^{(k)}) - \mathrm{math}(\boldsymbol{b}^{(k)})) \odot S$. In these relations, $\nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)})$ and $\nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)})$ denote the components of the gradient of minus the log-likelihood function computed at the point $\boldsymbol{\theta}^{(k)}$.

Jacobian of the gradient of the local model We use the general relation (E.5) to determine the tangent map of $\nabla_{\boldsymbol{\theta}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^N \times \mathbb{R}^N$:

$$\begin{aligned}T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{a}}\tilde{f}^{(k)}(\boldsymbol{\delta\theta}) =& H_{\boldsymbol{a}}^{(k)}\boldsymbol{\delta a} + L_1^1\ \mathrm{math}^*\left(\mathrm{math}(\boldsymbol{a})^{-1}\ \mathrm{math}(\boldsymbol{\delta a})\ \mathrm{math}(\boldsymbol{a})^{-1}\right) \\ &+ L_1^3\ \mathrm{math}^*((Z^{-1}(\mathrm{math}(\boldsymbol{\delta a}+\boldsymbol{\delta b}) \odot S)Z^{-1}) \odot S), \end{aligned} \tag{E.15}$$

$$\begin{aligned}T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{b}}\tilde{f}^{(k)}(\boldsymbol{\delta\theta}) =& H_{\boldsymbol{b}}^{(k)}\boldsymbol{\delta b} + L_1^2\ \mathrm{math}^*\left(\mathrm{math}(\boldsymbol{b})^{-1}\ \mathrm{math}(\boldsymbol{\delta b})\ \mathrm{math}(\boldsymbol{b})^{-1}\right) \\ &+ L_1^3\ \mathrm{math}^*((Z^{-1}(\mathrm{math}(\boldsymbol{\delta a}+\boldsymbol{\delta b}) \odot S)Z^{-1}) \odot S). \end{aligned} \tag{E.16}$$

The $(1,1)$ and $(1,2)$ blocks of the Jacobian matrix can be obtained from (E.15) by taking increments $\boldsymbol{\delta\theta}$ of the form $(\boldsymbol{\delta a}, \mathbf{0})$ and $(\mathbf{0}, \boldsymbol{\delta b})$, respectively. Analogously, the $(2,1)$ and $(2,2)$ blocks of the Jacobian matrix follow from (E.16) by using increments $\boldsymbol{\delta\theta}$ of the form $(\boldsymbol{\delta a}, \mathbf{0})$ and $(\mathbf{0}, \boldsymbol{\delta b})$, respectively.

Remark E.1 The non-targeted Hadamard DCC model contains an additional parameter matrix $C \in \mathbb{S}_n$ that is intrinsically parametrized with the vector $\boldsymbol{c} \in \mathbb{R}^N$, $N = \frac{1}{2}n(n+1)$, by setting $C := \mathrm{math}(\boldsymbol{c})$; the intrinsic parameters of the model are hence given by $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \in \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N$. The constraints of the model are listed in (2.6)-(2.7) and comprise three positive semidefiniteness/definiteness and two linear positivity conditions. More specifically, the conditions $M_1(\boldsymbol{a})$ and $M_2(\boldsymbol{a})$ necessary for the targeted Hadamard DCC model remain valid and, in addition, the three following constraints are imposed:

$$M_4(\boldsymbol{c}) := \mathrm{math}(\boldsymbol{c}) \succ 0 \in \mathbb{S}_n, \tag{E.17}$$

$$L_j(\boldsymbol{a}, \boldsymbol{b}) := \mathbf{i}_N - C_{\boldsymbol{a}}^{(j)}\boldsymbol{a} - C_{\boldsymbol{b}}^{(j)}\boldsymbol{b} > \mathbf{0}_N \in \mathbb{R}^N, \quad j = 1, 2, \tag{E.18}$$

where $C_{\boldsymbol{a}}^{(j)} = C_{\boldsymbol{b}}^{(j)} = (-1)^{j-1}\mathbb{I}_N$, $j = 1, 2$ with $\mathbb{I}_N$ the identity matrix. The local model corresponding to (3.7) is given in this case by (E.9) where the term that is associated to the constraint $M_3(\boldsymbol{a})$ is replaced by $L_1^3 D_{M_4}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, and where we add the term $\sum_{j=1}^2 L_2^j \mathbf{i}_N^\top D_{L_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ corresponding to (E.17) and (E.18); L_1^3 and $\mathbf{L}_2 := \left(L_2^1, L_2^2\right)^\top$ denote the penalization strengths. The log-likelihood function, its gradient, and the Jacobian of the gradient of the local model are obtained directly from (E.13)-(E.16) by taking into account the Remarks D.2 and D.4. The Bregman divergences associated to each of the constraints (E.6),(E.7), (E.8),(E.18) are obtained from (3.5) and (3.6):

$$D_{M_4}(\boldsymbol{c}, \boldsymbol{c}^{(k)}) = \mathrm{tr}(M_4(\boldsymbol{c}) \cdot M_4(\boldsymbol{c}^{(k)})^{-1}) - \log\det(M_4(\boldsymbol{c}) \cdot M_4(\boldsymbol{c}^{(k)})^{-1}) - n, \tag{E.19}$$

$$D_{L_j}^i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{1 + (-1)^j(a_i + b_i)}{1 + (-1)^j(a_i^{(k)} + b_i^{(k)})} - \log\frac{1 + (-1)^j(a_i + b_i)}{1 + (-1)^j(a_i^{(k)} + b_i^{(k)})} - 1, \tag{E.20}$$

with $i \in \{1,\dots,N\}$, $j = 1,2$. In each component of the gradient of the local model $\nabla_{\boldsymbol{a}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ and $\nabla_{\boldsymbol{b}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ in (E.13) and (E.14), respectively, the last summand has to be replaced by:

$$L_2^1 \sum_{i=1}^{N} \left[\frac{1}{1-a_i-b_i} - \frac{1}{1-a_i^{(k)}-b_i^{(k)}} \right] \cdot \mathbf{e}_N^{(i)} - L_2^2 \sum_{i=1}^{N} \left[\frac{1}{1+a_i+b_i} - \frac{1}{1+a_i^{(k)}+b_i^{(k)}} \right] \cdot \mathbf{e}_N^{(i)}, \tag{E.21}$$

with $\mathbf{e}_N^{(i)} \in \mathbb{R}^N$, $i \in \{1,\dots,N\}$ the canonical unit vectors introduced in Subsection A.1. Additionally, the third component of the gradient of the local model is given by

$$\nabla_{\boldsymbol{c}} \tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{c}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{c}}^{(k)}(\boldsymbol{c} - \boldsymbol{c}^{(k)}) - L_1^3 \operatorname{math}^*(\operatorname{math}(\boldsymbol{c})^{-1} - \operatorname{math}(\boldsymbol{c}^{(k)})^{-1}). \tag{E.22}$$

Concerning the Jacobian of the gradient of the local model $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}, \boldsymbol{\delta c})$, we use (E.15)-(E.16) where, again, the last summands are replaced by:

$$L_2^1 \sum_{i=1}^{N} \frac{\delta a_i + \delta b_i}{(1-a_i-b_i)^2} \cdot \mathbf{e}_N^{(i)} + L_2^2 \sum_{i=1}^{N} \frac{\delta a_i + \delta b_i}{(1+a_i+b_i)^2} \cdot \mathbf{e}_N^{(i)}, \tag{E.23}$$

and, additionally,

$$T_{\boldsymbol{\theta}} \nabla_{\boldsymbol{c}} \tilde{f}^{(k)}(\boldsymbol{\delta\theta}) = H_{\boldsymbol{c}}^{(k)} \boldsymbol{\delta c} + L_1^3 \operatorname{math}^* \left(\operatorname{math}(\boldsymbol{c})^{-1} \operatorname{math}(\boldsymbol{\delta c}) \operatorname{math}(\boldsymbol{c})^{-1} \right). \tag{E.24}$$

The relevant blocks of the Jacobian matrix are obtained from (E.15), (E.16) with (E.23) and from (E.24) by taking the corresponding increments $\boldsymbol{\delta\theta}$. For example, the elements $(1,1)$, $(1,2)$, and $(1,3)$ of the Jacobian require taking increments of the form $(\boldsymbol{\delta a}, \mathbf{0}, \mathbf{0})$, $(\mathbf{0}, \boldsymbol{\delta b}, \mathbf{0})$, and $(\mathbf{0}, \mathbf{0}, \boldsymbol{\delta c})$, respectively.

E.2 The rank deficient DCC model

As explained in Section 2.3, in this case the parameter matrices $A, B \in \mathbb{S}_n$ have a common prescribed rank $r \in \{1,\dots,n-1\}$, that is, $\operatorname{rank}(A) = \operatorname{rank}(B) = r$. The intrinsic parameterization is provided by two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{N^*}$, $N^* = nr - \frac{1}{2}r(r-1)$, using the operator $\operatorname{mat}_r : \mathbb{R}^{N^*} \longrightarrow \mathbb{L}_{n,r}$, presented in Subsection A.2 and setting: $A := \operatorname{mat}_r(\boldsymbol{a}) \operatorname{mat}_r(\boldsymbol{a})^\top$, $B := \operatorname{mat}_r(\boldsymbol{b}) \operatorname{mat}_r(\boldsymbol{b})^\top$. Let $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$; the intrinsic parameter subspace $\mathcal{P}$ is $\mathbb{R}^{N^*}$. The dynamics of the conditional correlation matrix process $\{Q_t\}$ for the rank deficient family is given in (2.18) and $\boldsymbol{\theta}$ is subjected to the identification and positivity constraints (2.19)-(2.20).

Constraints and the local model We group these constraints according to the classification in Subsection 3.2.2 which yields two linear and one positive definiteness constraint. In order to specify them let $\{i_1,\dots,i_r\}$ with $i_j = n(j-1) + \frac{1}{2}j(3-j)$, $j \in \{1,\dots,r\}$, be the entries spelled out in (A.12) of any vector in $\mathbb{R}^{N^*}$ that amount to the main diagonal of the corresponding matrix in $\mathbb{L}_{n,r}$ via the mat_r representation. Let $C_{\boldsymbol{a}}^{(1)}, C_{\boldsymbol{b}}^{(2)} \in \mathbb{M}_{r,N^*}$ be the matrices that have as rows the canonical unit vectors $\{\mathbf{e}_{N^*}^{(i_1)},\dots,\mathbf{e}_{N^*}^{(i_r)}\}$ with a minus sign in the front. Then:

$$L_1(\boldsymbol{a}) := -C_{\boldsymbol{a}}^{(1)} \boldsymbol{a} > \mathbf{0}_r, \tag{E.25}$$

$$L_2(\boldsymbol{b}) := -C_{\boldsymbol{b}}^{(2)} \boldsymbol{b} > \mathbf{0}_r, \tag{E.26}$$

$$M(\boldsymbol{\theta}) := (\mathbf{i}_n \mathbf{i}_n^\top - \operatorname{mat}_r(\boldsymbol{a}) \operatorname{mat}_r(\boldsymbol{a})^\top - \operatorname{mat}_r(\boldsymbol{b}) \operatorname{mat}_r(\boldsymbol{b})^\top) \odot S \succ 0, \tag{E.27}$$

where $L_1(\boldsymbol{a}), L_2(\boldsymbol{b}) \in \mathbb{R}^r$, and $M(\boldsymbol{\theta}) \in \mathbb{S}_n$. The local penalized model for the optimization problem (3.7) in this case is given by

$$\begin{aligned}\tilde{f}^{(k)}(\boldsymbol{\theta}) =& f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) \\ &+ \sum_{j=1}^{2} L_1^j \mathbf{i}_r^\top D_{L_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + L_2 D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}),\end{aligned} \tag{E.28}$$

where $\mathbf{L}_1 := \left(L_1^1, L_1^2\right)^\top$, L_2 specify the penalization strengths, $f(\boldsymbol{\theta}^{(k)})$ is minus the log-likelihood function evaluated at (3.1), $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})$ is its gradient, that is $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}^{(k)})$ which is determined by relations (D.1)-(D.5), and (D.7) in Proposition D.1. Finally, $H^{(k)}$ is its Hessian computed at the point $\boldsymbol{\theta}^{(k)}$. In relation (E.28) the Bregman divergences associated to each of the constraints (E.25)-(E.27) are easily obtained from (3.5)-(3.6):

$$D_{L_1}^j(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{a_{i_j}}{a_{i_j}^{(k)}} - \log \frac{a_{i_j}}{a_{i_j}^{(k)}} - 1, \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1, \ldots, r\}, \tag{E.29}$$

$$D_{L_2}^j(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{b_{i_j}}{b_{i_j}^{(k)}} - \log \frac{b_{i_j}}{b_{i_j}^{(k)}} - 1, \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1, \ldots, r\}, \tag{E.30}$$

$$D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \operatorname{tr}(M(\boldsymbol{\theta}) \cdot M(\boldsymbol{\theta}^{(k)})^{-1}) - \log \det(M(\boldsymbol{\theta}) \cdot M(\boldsymbol{\theta}^{(k)})^{-1}) - n. \tag{E.31}$$

Gradient of the local model A straightforward computation provided in Technical Appendix F.2 provides the following expressions for the components of the gradient of the local model (E.28):

$$\begin{aligned}\nabla_{\boldsymbol{a}} \tilde{f}^{(k)}(\boldsymbol{\theta}) =& \nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{a}}^{(k)}(\boldsymbol{a} - \boldsymbol{a}^{(k)}) - L_1^1 \sum_{j=1}^{r} \left[\frac{1}{a_{i_j}} - \frac{1}{a_{i_j}^{(k)}}\right] \cdot \mathbf{e}_{N^*}^{(i_j)} \\ &+ 2L_2 \operatorname{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \operatorname{mat}_r(\boldsymbol{a})),\end{aligned} \tag{E.32}$$

$$\begin{aligned}\nabla_{\boldsymbol{b}} \tilde{f}^{(k)}(\boldsymbol{\theta}) =& \nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{b}}^{(k)}(\boldsymbol{b} - \boldsymbol{b}^{(k)}) - L_1^2 \sum_{j=1}^{r} \left[\frac{1}{b_{i_j}} - \frac{1}{b_{i_j}^{(k)}}\right] \cdot \mathbf{e}_{N^*}^{(i_j)} \\ &+ 2L_2 \operatorname{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \operatorname{mat}_r(\boldsymbol{b})),\end{aligned} \tag{E.33}$$

where $i_j = n(j-1) + \frac{1}{2}j(3-j)$, $j \in \{1, \ldots, r\}$, $\nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)})$ and $\nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)})$ are the components of the gradient of minus the log-likelihood function computed at the point $\boldsymbol{\theta}^{(k)}$; $\mathbf{e}_{N^*}^{(i)} \in \mathbb{R}^{N^*}$, $i \in \{1, \ldots, N^*\}$ are the canonical unit vectors.

Jacobian of the gradient of the local model We use the general relation (E.5) to determine the tangent map to $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$:

$$\begin{aligned}T_{\boldsymbol{\theta}} \nabla_{\boldsymbol{a}} \tilde{f}^{(k)}(\boldsymbol{\delta a}, \boldsymbol{\delta b}) =& H_{\boldsymbol{a}}^{(k)} \boldsymbol{\delta a} + L_1^1 \sum_{j=1}^{r} \frac{\delta a_{i_j}}{a_{i_j}^2} \cdot \mathbf{e}_{N^*}^{(i_j)} + 2L_2 \operatorname{vec}_r(((M(\boldsymbol{\theta})^{-1}(W \odot S) M(\boldsymbol{\theta})^{-1}) \odot S) \operatorname{mat}_r(\boldsymbol{a}) \\ &+ ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \operatorname{mat}_r(\boldsymbol{\delta a})),\end{aligned} \tag{E.34}$$

$$\begin{aligned}T_{\boldsymbol{\theta}} \nabla_{\boldsymbol{b}} \tilde{f}^{(k)}(\boldsymbol{\delta a}, \boldsymbol{\delta b}) =& H_{\boldsymbol{b}}^{(k)} \boldsymbol{\delta b} + L_1^2 \sum_{j=1}^{r} \frac{\delta b_{i_j}}{b_{i_j}^2} \cdot \mathbf{e}_{N^*}^{(i_j)} + 2L_2 \operatorname{vec}_r(((M(\boldsymbol{\theta})^{-1}(W \odot S) M(\boldsymbol{\theta})^{-1}) \odot S) \operatorname{mat}_r(\boldsymbol{b}) \\ &+ ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \operatorname{mat}_r(\boldsymbol{\delta b})),\end{aligned} \tag{E.35}$$

with $i_j = n(j-1) + \frac{1}{2}j(3-j)$, $j \in \{1,\dots,r\}$, $W := \mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top + \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top$.

The (1,1) and (1, 2) blocks of the Jacobian matrix can be obtained from (E.34) by taking $\boldsymbol{\delta\theta}$ of the form $(\boldsymbol{\delta a}, \mathbf{0})$, $(\mathbf{0}, \boldsymbol{\delta b})$. Analogously, the (2,1) and (2, 2) blocks of the Jacobian are prescribed by (E.35) by setting $\boldsymbol{\delta\theta}$ as $(\boldsymbol{\delta a}, \mathbf{0})$, $(\mathbf{0}, \boldsymbol{\delta b})$, respectively.

Remark E.2 The intrinsic parameters $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \times \mathbb{R}^N$, $N := \frac{1}{2}n(n+1)$ of the non-targeted rank deficient DCC models are subjected to the identification, positivity, and stationarity constraints in (2.21), (2.22), and (2.23), respectively. According to the classification in Subsection 3.2.2 they amount to three linear and two non-linear constraints. The first two linear constraints are provided in (E.25) and (E.26). In order to specify the others, we use the same notation introduced earlier:

$$M_1(\boldsymbol{c}) := \mathrm{math}(\boldsymbol{c}) \succ 0, \tag{E.36}$$

$$N_1(\boldsymbol{\theta}) := \mathbf{i}_N - \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{a})\,\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{b})\,\mathrm{mat}_r(\boldsymbol{b})^\top) > \mathbf{0}_N, \tag{E.37}$$

$$N_2(\boldsymbol{\theta}) := \mathbf{i}_N + \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{a})\,\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{b})\,\mathrm{mat}_r(\boldsymbol{b})^\top) > \mathbf{0}_N, \tag{E.38}$$

where $M_1(\boldsymbol{c}) \in \mathbb{S}_n$ and $N_1(\boldsymbol{\theta}), N_2(\boldsymbol{\theta}) \in \mathbb{R}^N$. The local penalized model for the optimization problem (3.7) in this case is obtained from (E.28) by replacing the last summand associated to the constraint (E.27) by $L_2 D_{M_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^2 L_3^j \mathbf{i}_N^\top D_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, with L_2 and the vector $\mathbf{L}_3 := \left(L_3^1, L_3^2\right)^\top$ the penalization strengths. The log-likelihood function, its gradient, and the Jacobian of the gradient of the local model are obtained directly from (D.1)-(D.5), and (D.7) in Proposition D.1 taking into account Remarks D.2 and D.4. In relation (E.28) the Bregman divergences associated to the additional constraints (E.36)-(E.38) are easily obtained from (3.5) and (3.6), respectively:

$$D_{M_1}(\boldsymbol{c}, \boldsymbol{c}^{(k)}) = \mathrm{tr}(M_1(\boldsymbol{c}) \cdot M_1(\boldsymbol{c}^{(k)})^{-1}) - \log\det(M_1(\boldsymbol{c}) \cdot M_1(\boldsymbol{c}^{(k)})^{-1}) - n, \tag{E.39}$$

$$D^i_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{(N_j(\boldsymbol{\theta}))_i}{(N_j(\boldsymbol{\theta}^{(k)}))_i} - \log\frac{(N_1(\boldsymbol{\theta}))_i}{(N_j(\boldsymbol{\theta}^{(k)}))_i} - 1, \;\; i \in \{1,\dots,N\}, \;\; j = \{1,2\}. \tag{E.40}$$

The last summand in the expression (E.32) of the component $\nabla_{\boldsymbol{a}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ of the gradient of the local model has to be replaced by:

$$2\sum_{i=1}^N \left(L_3^1 \left[\frac{1}{(N_1(\boldsymbol{\theta}))_i} - \frac{1}{(N_1(\boldsymbol{\theta}^{(k)}))_i} \right] - L_3^2 \left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i} \right] \right) \cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{a})),$$

analogously, in the expression (E.33) of $\nabla_{\boldsymbol{b}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ by

$$2\sum_{i=1}^N \left(L_3^1 \left[\frac{1}{(N_1(\boldsymbol{\theta}))_i} - \frac{1}{(N_1(\boldsymbol{\theta}^{(k)}))_i} \right] - L_3^2 \left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i} \right] \right) \cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{b})),$$

and, finally, the component $\nabla_{\boldsymbol{c}}\tilde{f}^{(k)}(\boldsymbol{\theta})$ is written as

$$\nabla_{\boldsymbol{c}}\tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{c}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{c}}^{(k)}(\boldsymbol{c} - \boldsymbol{c}^{(k)}) - L_2\, \mathrm{math}^*(\mathrm{math}(\boldsymbol{c})^{-1} - \mathrm{math}(\boldsymbol{c}^{(k)})^{-1}). \tag{E.41}$$

Concerning the Jacobian of the gradient of the local model, we consider $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}, \boldsymbol{\delta c}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \times \mathbb{R}^N$ and use (E.34)-(E.35) where, again, the last summands are replaced by:

$$2\sum_{i=1}^N \left\{ \left(L_3^1 \left[\frac{1}{(N_1(\boldsymbol{\theta}))_i} - \frac{1}{(N_1(\boldsymbol{\theta}^{(k)}))_i} \right] - L_3^2 \left[\frac{1}{(N_2(\boldsymbol{\theta}))_i} - \frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i} \right] \right) \cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{\delta a})) \right.$$
$$\left. + (\mathrm{vech}(W))_i \cdot \left(\frac{L_3^1}{(N_1(\boldsymbol{\theta}))_i^2} + \frac{L_3^2}{(N_2(\boldsymbol{\theta}))_i^2} \right) \cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{mat}_r(\boldsymbol{a})) \right\} \tag{E.42}$$

and by

$$2\sum_{i=1}^{N}\left\{\left(L_3^1\left[\frac{1}{(N_1(\boldsymbol{\theta}))_i}-\frac{1}{(N_1(\boldsymbol{\theta}^{(k)}))_i}\right]-L_3^2\left[\frac{1}{(N_2(\boldsymbol{\theta}))_i}-\frac{1}{(N_2(\boldsymbol{\theta}^{(k)}))_i}\right]\right)\cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot \mathrm{mat}_r(\boldsymbol{\delta b}))\right.$$
$$\left.+(\mathrm{vech}(W))_i\cdot\left(\frac{L_3^1}{(N_1(\boldsymbol{\theta}))_i^2}+\frac{L_3^2}{(N_2(\boldsymbol{\theta}))_i^2}\right)\cdot \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot \mathrm{mat}_r(\boldsymbol{b}))\right\}, \tag{E.43}$$

respectively. In these expressions $W := \mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top + \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top$. Additionally we have

$$T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{c}}\tilde{f}^{(k)}(\boldsymbol{\delta\theta}) = H_{\boldsymbol{c}}^{(k)}\boldsymbol{\delta c} + L_2\ \mathrm{math}^*\left(\mathrm{math}(\boldsymbol{c})^{-1}\ \mathrm{math}(\boldsymbol{\delta c})\ \mathrm{math}(\boldsymbol{c})^{-1}\right). \tag{E.44}$$

The relevant blocks of the Jacobian matrix are obtained from (E.34) and (E.35) by taking into account (E.42) and (E.43), respectively, and from (E.44) by taking the corresponding increments $\boldsymbol{\delta\theta}$. For example, the elements $(1,1)$, $(1,2)$, and $(1,3)$ of the Jacobian are obtained by taking increments of the form $(\boldsymbol{\delta a},\mathbf{0},\mathbf{0})$, $(\mathbf{0},\boldsymbol{\delta b},\mathbf{0})$, and $(\mathbf{0},\mathbf{0},\boldsymbol{\delta c})$, respectively.

E.3 The Almon DCC model

The Almon DCC model introduced in Section 2.4 can be seen as a particular case of the rank deficient DCC model with parameter matrices $A, B \in \mathbb{S}_n$ in (2.8) of rank $r = 1$. The matrices are intrinsically parametrized by two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$ via the Almon lag operator $\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$ (see Subsection A.2 for the definition and properties of this operator) by defining $A := \mathrm{alm}_n(\boldsymbol{a})\ \mathrm{alm}_n(\boldsymbol{a})^\top$ and $B := \mathrm{alm}_n(\boldsymbol{b})\ \mathrm{alm}_n(\boldsymbol{b})^\top$. Let $\boldsymbol{\theta} := (\boldsymbol{a},\boldsymbol{b}) \in \mathbb{R}^3 \times \mathbb{R}^3$; in this case the intrinsic parameter subspace $\mathcal{P}$ is $\mathbb{R}^3$ of dimension $P = 3$. The DCC model in the Almon specification is spelled out in detail in (2.24). The parameter constraints that are imposed in order to ensure that the process admits a stationary solution and that the resulting conditional correlation and hence the conditional covariance matrices are positive definite, are given by the relations (2.25)-(2.26).

Constraints and the local model We group the parameter constraints following the classification introduced in Subsection 3.2.2. This leads to the following set of two nonlinear positivity and one positive definiteness constraint:

$$N_1(\boldsymbol{a}) := (\mathrm{alm}_n(\boldsymbol{a}))_1 = \mathbf{e}_n^{(1)\top}\cdot \mathrm{alm}_n(\boldsymbol{a}) > 0, \tag{E.45}$$
$$N_2(\boldsymbol{b}) := (\mathrm{alm}_n(\boldsymbol{b}))_1 = \mathbf{e}_n^{(1)\top}\cdot \mathrm{alm}_n(\boldsymbol{b}) > 0, \tag{E.46}$$
$$M(\boldsymbol{\theta}) := (\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{alm}_n(\boldsymbol{a})\,\mathrm{alm}_n(\boldsymbol{a})^\top - \mathrm{alm}_n(\boldsymbol{b})\,\mathrm{alm}_n(\boldsymbol{b})^\top)\odot S \succ 0, \tag{E.47}$$

where $N_1(\boldsymbol{a}), N_2(\boldsymbol{b}) \in \mathbb{R}$, $M(\boldsymbol{\theta}) \in \mathbb{S}_n$, and $\mathbf{e}_n^{(1)} \in \mathbb{R}^n$ is the canonical unit vector. In the case of the Almon family of models, the local penalized model for the optimization problem (3.7) with the Bregman divergences associated to the constraints (E.45)-(E.47) is given by

$$\tilde{f}^{(k)}(\boldsymbol{\theta}) = f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)})^\top H^{(k)}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)})$$
$$+\sum_{j=1}^{2} L_1^j \mathbf{i}_{m_j}^\top D_{N_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) + L_2 D_M(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}), \tag{E.48}$$

where $\mathbf{L}_2 := \left(L_1^1, L_1^2\right)^\top$ and L_2 specify the penalization strengths and $m_j = \dim\left\{D_{N_j}(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})\right\}$, $j \in \{1,2\}$. In this relation $f(\boldsymbol{\theta}^{(k)})$ is minus the log-likelihood function evaluated at (3.1), $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})$ is its

gradient, that is $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}^{(k)})$, which is determined by relations (D.1)-(D.5), and (D.8) in Proposition D.1; $H^{(k)}$ is the Hessian of the minus log-likelihood computed at the point $\boldsymbol{\theta}^{(k)}$; $D_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, $j = \{1, 2\}$ and $D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ are the Bregman divergences associated to the constraints (E.45)-(E.46) and (E.47), respectively. By (3.6) and (3.5), the expressions of these divergences are

$$D_{N_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{(\mathrm{alm}_n(\boldsymbol{a}))_1}{(\mathrm{alm}_n(\boldsymbol{a})^{(k)})_1} - \log \frac{(\mathrm{alm}_n(\boldsymbol{a}))_1}{(\mathrm{alm}_n(\boldsymbol{a})^{(k)})_1} - 1, \tag{E.49}$$

$$D_{N_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{(\mathrm{alm}_n(\boldsymbol{b}))_1}{(\mathrm{alm}_n(\boldsymbol{b})^{(k)})_1} - \log \frac{(\mathrm{alm}_n(\boldsymbol{b}))_1}{(\mathrm{alm}_n(\boldsymbol{b})^{(k)})_1} - 1, \tag{E.50}$$

$$D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \mathrm{tr}(M(\boldsymbol{a}, \boldsymbol{b}) \cdot M(\boldsymbol{a}^{(k)}, \boldsymbol{b}^{(k)})^{-1}) - \log \det(M(\boldsymbol{a}, \boldsymbol{b}) \cdot M(\boldsymbol{a}^{(k)}, \boldsymbol{b}^{(k)})^{-1}) - n. \tag{E.51}$$

Gradient of the local model A computation that is explicitly provided in Technical Appendix F.3 provides the following expressions for the components of the gradient of the local model (E.48):

$$\begin{aligned}\nabla_{\boldsymbol{a}} \tilde{f}^{(k)}(\boldsymbol{\theta}) =& \nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{a}}^{(k)}(\boldsymbol{a} - \boldsymbol{a}^{(k)}) - L_1^1 \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{a}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{a})^{(k)})_1}\right] \left(K_{\boldsymbol{a}}^\top \cdot \mathbf{e}_n^{(1)}\right) \\ &+ 2L_2 K_{\boldsymbol{a}}^\top ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{a}),\end{aligned} \tag{E.52}$$

$$\begin{aligned}\nabla_{\boldsymbol{b}} \tilde{f}^{(k)}(\boldsymbol{\theta}) =& \nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{b}}^{(k)}(\boldsymbol{b} - \boldsymbol{b}^{(k)}) - L_1^2 \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{b}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{b})^{(k)})_1}\right] \left(K_{\boldsymbol{b}}^\top \cdot \mathbf{e}_n^{(1)}\right) \\ &+ 2L_2 K_{\boldsymbol{b}}^\top ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{b}),\end{aligned} \tag{E.53}$$

where $\nabla_{\boldsymbol{a}} f(\boldsymbol{\theta}^{(k)})$ and $\nabla_{\boldsymbol{b}} f(\boldsymbol{\theta}^{(k)})$ are the components of the gradient of minus the log-likelihood function computed at the point $\boldsymbol{\theta}^{(k)}$. Recall that $K_{\boldsymbol{a}}, K_{\boldsymbol{b}} \in \mathbb{M}_{n,3}$ are the matrices introduced in (A.16).

Jacobian of the gradient of the local model We use the expression (E.5) to determine the tangent map to $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^3 \times \mathbb{R}^3$. Computations detailed in Technical Appendix F.3 provide the following results:

$$\begin{aligned}&T_{\boldsymbol{\theta}} \nabla_{\boldsymbol{a}} \tilde{f}^{(k)}(\boldsymbol{\delta a}, \boldsymbol{\delta b}) = \\ &H_{\boldsymbol{a}}^{(k)} \boldsymbol{\delta a} + L_1^1 \left\{ \frac{(K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})_1}{(\mathrm{alm}_n(\boldsymbol{a}))_1^2} \cdot (K_{\boldsymbol{a}}^\top \cdot \mathbf{e}_n^{(1)}) - \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{a}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{a}^{(k)}))_1}\right] ((K_{\boldsymbol{aa}} \cdot \boldsymbol{\delta a})^\top \cdot \mathbf{e}_n^{(1)}) \right\} \\ &+ 2L_2 \big\{ (K_{\boldsymbol{aa}} \cdot \boldsymbol{\delta a})^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{a}) \\ &+ K_{\boldsymbol{a}} \cdot ((M(\boldsymbol{\theta})^{-1}(W \odot S) M(\boldsymbol{\theta})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{a}) + K_{\boldsymbol{a}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \big\}\end{aligned} \tag{E.54}$$

$$\begin{aligned}&T_{\boldsymbol{\theta}} \nabla_{\boldsymbol{b}} \tilde{f}^{(k)}(\boldsymbol{\delta a}, \boldsymbol{\delta b}) = \\ &H_{\boldsymbol{b}}^{(k)} \boldsymbol{\delta b} + L_1^2 \left\{ \frac{(K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})_1}{(\mathrm{alm}_n(\boldsymbol{b}))_1^2} \cdot (K_{\boldsymbol{b}}^\top \cdot \mathbf{e}_n^{(1)}) - \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{b}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{b}^{(k)}))_1}\right] ((K_{\boldsymbol{bb}} \cdot \boldsymbol{\delta b})^\top \cdot \mathbf{e}_n^{(1)}) \right\} \\ &+ 2L_2 \big\{ (K_{\boldsymbol{bb}} \cdot \boldsymbol{\delta b})^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{b}) \\ &+ K_{\boldsymbol{b}} \cdot ((M(\boldsymbol{\theta})^{-1}(W \odot S) M(\boldsymbol{\theta})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{b}) + K_{\boldsymbol{b}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \big\}\end{aligned} \tag{E.55}$$

with

$$W := (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top + \mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top + (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^\top + \mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top,$$
$$K_{\boldsymbol{aa}} \cdot \boldsymbol{\delta a} := (\mathbf{0}_n \mid \mathbf{k}_n^1 \odot (K_{\boldsymbol{a}}^0 \cdot \boldsymbol{\delta a}) \mid \mathbf{k}_n^2 \odot (K_{\boldsymbol{a}}^0 \cdot \boldsymbol{\delta a})),$$
$$K_{\boldsymbol{bb}} \cdot \boldsymbol{\delta b} := (\mathbf{0}_n \mid \mathbf{k}_n^1 \odot (K_{\boldsymbol{b}}^0 \cdot \boldsymbol{\delta b}) \mid \mathbf{k}_n^2 \odot (K_{\boldsymbol{b}}^0 \cdot \boldsymbol{\delta b})),$$
$$K_{\boldsymbol{a}}^0 := \left(\mathbf{0}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\boldsymbol{a}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\boldsymbol{a}})\right),$$
$$K_{\boldsymbol{b}}^0 := \left(\mathbf{0}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\boldsymbol{b}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\boldsymbol{b}})\right).$$

Despite their apparent complexity, the expressions in (E.54)-(E.55) are explicit and allow the computation of the blocks of the Jacobian matrix. The (1,1) and (1, 2) blocks of the Jacobian matrix can be obtained from (E.54) by taking increments $\boldsymbol{\delta\theta}$ of the form $(\boldsymbol{\delta a}, \mathbf{0})$, $(\mathbf{0}, \boldsymbol{\delta b})$. Analogously, the (2,1) and (2, 2) blocks of the Jacobian are obtained out of (E.55) by setting $\boldsymbol{\delta\theta}$ as $(\boldsymbol{\delta a}, \mathbf{0})$, $(\mathbf{0}, \boldsymbol{\delta b})$, respectively.

Remark E.3 The intrinsic parameters $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^N$, $N := \frac{1}{2}n(n+1)$ of the non-targeted version of the Almon DCC model are subjected to the identification, positivity, and stationarity constraints (2.27), (2.28), and (2.29), respectively. According to the classification in Subsection 3.2.2 they amount to four nonlinear constraints and one positive semidefiniteness constraint. Two of the four nonlinear constraints are provided in (E.45) and (E.46), and the other conditions are given by:

$$N_3(\boldsymbol{\theta}) := \mathbf{i}_N - \mathrm{vech}(\mathrm{alm}_n(\boldsymbol{a})\,\mathrm{alm}_n(\boldsymbol{a})^\top + \mathrm{alm}_n(\boldsymbol{b})\,\mathrm{alm}_n(\boldsymbol{b})^\top) > \mathbf{0}_N, \tag{E.56}$$
$$N_4(\boldsymbol{\theta}) := \mathbf{i}_N + \mathrm{vech}(\mathrm{alm}_n(\boldsymbol{a})\,\mathrm{alm}_n(\boldsymbol{a})^\top + \mathrm{alm}_n(\boldsymbol{b})\,\mathrm{alm}_n(\boldsymbol{b})^\top) > \mathbf{0}_N, \tag{E.57}$$
$$M_1(\boldsymbol{c}) := \mathrm{math}(\boldsymbol{c}) \succ 0, \tag{E.58}$$

where $N_3(\boldsymbol{\theta}), N_4(\boldsymbol{\theta}) \in \mathbb{R}^N$, and $M_1(\boldsymbol{c}) \in \mathbb{S}^n$.

The local penalized model for the optimization problem (3.7) in this case is obtained from (E.48) by replacing the last summand associated to the constraint (E.47) by $\sum_{j=3}^4 L_1^j \mathbf{i}_{m_j}^\top D_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) + L_2 D_{M_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, with the penalization strengths L_2 and L_1^3, L_1^4 the additional components of the vector $\mathbf{L}_1$ in (E.48). The log-likelihood function, its gradient, and the Jacobian of the gradient of the local model are obtained directly from (D.1)-(D.5), and (D.8) in Proposition D.1 by taking into account Remarks D.2 and D.4. In relation (E.48), the Bregman divergences associated to the additional constraints (E.56)-(E.57) are easily obtained from (3.6), and (E.58) from (3.5) in the following way:

$$D_{N_j}^i(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{(N_j(\boldsymbol{\theta}))_i}{(N_j(\boldsymbol{\theta}^{(k)}))_i} - \log\frac{(N_j(\boldsymbol{\theta}))_i}{(N_j(\boldsymbol{\theta}^{(k)}))_i} - 1, \;\; i \in \{1, \ldots, N\}, \;\; j \in \{3, 4\}. \tag{E.59}$$
$$D_{M_1}(\boldsymbol{c}, \boldsymbol{c}^{(k)}) = \mathrm{tr}(M_1(\boldsymbol{c}) \cdot M_1(\boldsymbol{c}^{(k)})^{-1}) - \log\,\det(M_1(\boldsymbol{c}) \cdot M_1(\boldsymbol{c}^{(k)})^{-1}) - n. \tag{E.60}$$

The last summand in the expression (E.52) of the component $\nabla_{\boldsymbol{a}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ of the gradient of the local model has to be replaced by

$$2\sum_{i=1}^N \left(L_1^3 \left[\frac{1}{(N_3(\boldsymbol{\theta}))_i} - \frac{1}{(N_3(\boldsymbol{\theta}^{(k)}))_i} \right] - L_1^4 \left[\frac{1}{(N_4(\boldsymbol{\theta}))_i} - \frac{1}{(N_4(\boldsymbol{\theta}^{(k)}))_i} \right] \right) (K_{\boldsymbol{a}} \cdot \mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{alm}_n(\boldsymbol{a})), \tag{E.61}$$

the same last summand in (E.53) becomes

$$2\sum_{i=1}^N \left(L_1^3 \left[\frac{1}{(N_3(\boldsymbol{\theta}))_i} - \frac{1}{(N_3(\boldsymbol{\theta}^{(k)}))_i} \right] - L_1^4 \left[\frac{1}{(N_4(\boldsymbol{\theta}))_i} - \frac{1}{(N_4(\boldsymbol{\theta}^{(k)}))_i} \right] \right) (K_{\boldsymbol{b}} \cdot \mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{alm}_n(\boldsymbol{b})), \tag{E.62}$$

and additionally we have:

$$\nabla_{\boldsymbol{c}}\tilde{f}^{(k)}\left(\boldsymbol{\theta}\right)=\nabla_{\boldsymbol{c}}f(\boldsymbol{\theta}^{(k)})+H_{\boldsymbol{c}}^{(k)}(\boldsymbol{c}-\boldsymbol{c}^{(k)})-L_2\ \mathrm{math}^*(\mathrm{math}(\boldsymbol{c})^{-1}-\mathrm{math}(\boldsymbol{c}^{(k)})^{-1}),\tag{E.63}$$

where $K_{\boldsymbol{a}},K_{\boldsymbol{b}}\in\mathbb{M}_{n,3}$ are the matrices introduced in (A.16) and $\mathbf{e}_N^{(i)}$, $i\in\{1,\ldots,N\}$, are the canonical unit vectors introduced in Subsection A.1. Concerning the Jacobian of the gradient of the local model, we consider an arbitrary increment $\boldsymbol{\delta\theta}:=(\boldsymbol{\delta a},\boldsymbol{\delta b},\boldsymbol{\delta c})\in\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}^N$ and use the expressions (E.54) and (E.55) where, again, the last summand is replaced by:

$$\begin{aligned}2\sum_{i=1}^{N}\Bigg\{&\left(L_1^3\left[\frac{1}{(N_3(\boldsymbol{\theta}))_i}-\frac{1}{(N_3(\boldsymbol{\theta}^{(k)}))_i}\right]-L_1^4\left[\frac{1}{(N_4(\boldsymbol{\theta}))_i}-\frac{1}{(N_4(\boldsymbol{\theta}^{(k)}))_i}\right]\right)\cdot((K_{\boldsymbol{aa}}\cdot\boldsymbol{\delta a})^\top\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot\mathrm{alm}_n(\boldsymbol{a})\\&+K_{\boldsymbol{a}}^\top\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot(K_{\boldsymbol{a}}\cdot\boldsymbol{\delta a}))+(\mathrm{vech}(W))_i\cdot\left[\frac{L_1^3}{(N_3(\boldsymbol{\theta}))_i^2}+\frac{L_1^4}{(N_4(\boldsymbol{\theta}))_i^2}\right]\cdot(K_{\boldsymbol{a}}^\top\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot\mathrm{alm}_n(\boldsymbol{a}))\Bigg\},\end{aligned}\tag{E.64}$$

and by

$$\begin{aligned}2\sum_{i=1}^{N}\Bigg\{&\left(L_1^3\left[\frac{1}{(N_3(\boldsymbol{\theta}))_i}-\frac{1}{(N_3(\boldsymbol{\theta}^{(k)}))_i}\right]-L_1^4\left[\frac{1}{(N_4(\boldsymbol{\theta}))_i}-\frac{1}{(N_4(\boldsymbol{\theta}^{(k)}))_i}\right]\right)\cdot((K_{\boldsymbol{bb}}\cdot\boldsymbol{\delta b})^\top\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot\mathrm{alm}_n(\boldsymbol{b})\\&+K_{\boldsymbol{b}}^\top\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot(K_{\boldsymbol{b}}\cdot\boldsymbol{\delta b}))+(\mathrm{vech}(W))_i\cdot\left[\frac{L_1^3}{(N_3(\boldsymbol{\theta}))_i^2}+\frac{L_1^4}{(N_4(\boldsymbol{\theta}))_i^2}\right]\cdot(K_{\boldsymbol{b}}^\top\cdot\mathrm{vech}^*(\mathbf{e}_N^{(i)})\cdot\mathrm{alm}_n(\boldsymbol{b}))\Bigg\},\end{aligned}\tag{E.65}$$

respectively, with $W:=(K_{\boldsymbol{a}}\cdot\boldsymbol{\delta a})\cdot\mathrm{alm}_n(\boldsymbol{a})^\top+\mathrm{alm}_n\left(\boldsymbol{a}\right)\cdot(K_{\boldsymbol{a}}\cdot\boldsymbol{\delta a})^\top+(K_{\boldsymbol{b}}\cdot\boldsymbol{\delta b})\cdot\mathrm{alm}_n\left(\boldsymbol{b}\right)^\top+\mathrm{alm}_n\left(\boldsymbol{b}\right)\cdot(K_{\boldsymbol{b}}\cdot\boldsymbol{\delta b})^\top$ and $K_{\boldsymbol{aa}}$, $K_{\boldsymbol{bb}}$ as in (E.54) and (E.55). Finally, we also have

$$T_{\boldsymbol{\theta}}\nabla_{\boldsymbol{c}}\tilde{f}^{(k)}\left(\boldsymbol{\delta\theta}\right)=H_{\boldsymbol{c}}^{(k)}\boldsymbol{\delta c}+L_2\ \mathrm{math}^*\left(\mathrm{math}(\boldsymbol{c})^{-1}\ \mathrm{math}(\boldsymbol{\delta c})\ \mathrm{math}(\boldsymbol{c})^{-1}\right).\tag{E.66}$$

The relevant blocks of the Jacobian matrix are obtained from (E.54), (E.55) with (E.64) and (E.65), respectively, and from (E.66) by taking the corresponding increments $\boldsymbol{\delta\theta}$. For example, the elements $(1,1)$, $(1,2)$, and $(1,3)$ of the Jacobian require taking increments of the form $(\boldsymbol{\delta a},\mathbf{0},\mathbf{0})$, $(\mathbf{0},\boldsymbol{\delta b},\mathbf{0})$, and $(\mathbf{0},\mathbf{0},\boldsymbol{\delta c})$, respectively.

E.4 The scalar DCC model

In this case the parameter matrices $A,B\in\mathbb{S}_n$ in (2.8) are of the form $A=a\mathbf{i}_n\mathbf{i}_n^\top$, $B=b\mathbf{i}_n\mathbf{i}_n^\top$, with $a,b\in\mathbb{R}$. The intrinsic parameter subspace $\mathcal{P}$ is in this case $\mathbb{R}$ and its dimension $P=1$; we denote $\boldsymbol{\theta}:=(a,b)\in\mathbb{R}\times\mathbb{R}$. The scalar DCC model specification together with the associated stationarity and positivity constraints are provided in (2.30)-(2.31).

Constraints and the local model In this case, the necessary constraints reduce to three linear positivity conditions, namely:

$$L_1(\boldsymbol{\theta}):=1-C_a^{(1)}a-C_b^{(1)}b>0,\tag{E.67}$$

$$L_2(a):=-C_a^{(2)}a\geq 0,\tag{E.68}$$

$$L_3(b):=-C_b^{(3)}b\geq 0,\tag{E.69}$$

where $L_1(\boldsymbol{\theta}), L_2(a), L_3(b) \in \mathbb{R}$, $C_a^{(1)} = C_b^{(1)} = 1$, and $C_a^{(2)} = C_b^{(3)} = -1$. The local penalized model for the optimization problem (3.7) is given in this case by

$$\tilde{f}^{(k)}(\boldsymbol{\theta}) = f(\boldsymbol{\theta}^{(k)}) + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})^{\top} H^{(k)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \sum_{j=1}^{3} L_1^j D_{L_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}), \quad \text{(E.70)}$$

where the vector $\mathbf{L}_1 := \left(L_1^1, L_1^2, L_1^3\right)^{\top}$ contains the penalization strengths, $f(\boldsymbol{\theta}^{(k)})$ is minus the log-likelihood function in (3.1), $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}^{(k)}$ is its gradient, determined by relations (D.1)-(D.5), and (D.9) in Proposition D.1, and $H^{(k)}$ is its Hessian computed at the point $\boldsymbol{\theta}^{(k)}$. The Bregman divergences associated to each of the constraints (E.67)-(E.69) and obtained from (3.6) take the following expressions:

$$D_{L_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{1-a-b}{1-a^{(k)}-b^{(k)}} - \log \frac{1-a-b}{1-a^{(k)}-b^{(k)}} - 1, \quad \text{(E.71)}$$

$$D_{L_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{a}{a^{(k)}} - \log \frac{a}{a^{(k)}} - 1, \quad \text{(E.72)}$$

$$D_{L_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \frac{b}{b^{(k)}} - \log \frac{b}{b^{(k)}} - 1. \quad \text{(E.73)}$$

Gradient of the local model A straightforward computation yields:

$$\nabla_a \tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_a f(\boldsymbol{\theta}^{(k)}) + H_a^{(k)}(a - a^{(k)}) + L_1^1 \left[\frac{1}{1-a-b} - \frac{1}{1-a^{(k)}-b^{(k)}}\right] - L_1^2 \left[\frac{1}{a} - \frac{1}{a^{(k)}}\right], \quad \text{(E.74)}$$

$$\nabla_b \tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_b f(\boldsymbol{\theta}^{(k)}) + H_b^{(k)}(b - b^{(k)}) + L_1^1 \left[\frac{1}{1-a-b} - \frac{1}{1-a^{(k)}-b^{(k)}}\right] - L_1^3 \left[\frac{1}{b} - \frac{1}{b^{(k)}}\right], \quad \text{(E.75)}$$

where $\nabla_a f(\boldsymbol{\theta}^{(k)})$, $\nabla_b f(\boldsymbol{\theta}^{(k)})$ are the components of the gradient of minus the log-likelihood function $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}^{(k)}) = -\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}^{(k)}, \mathbf{r})$ determined by the expressions (D.1)-(D.5) and by the relation (D.9).

Jacobian of the gradient of the local model A straightforward computation yields, for any $(\delta a, \delta b) \in \mathbb{R} \times \mathbb{R}$, the following expressions:

$$T_{\boldsymbol{\theta}} \nabla_a \tilde{f}^{(k)}(\delta a, \delta b) = H_a^{(k)} \delta a + L_1^1 \frac{\delta a + \delta b}{(1-a-b)^2} + L_1^2 \frac{\delta a}{a^2}, \quad \text{(E.76)}$$

$$T_{\boldsymbol{\theta}} \nabla_b \tilde{f}^{(k)}(\delta a, \delta b) = H_b^{(k)} \delta b + L_1^1 \frac{\delta b + \delta b}{(1-a-b)^2} + L_1^3 \frac{\delta b}{b^2}. \quad \text{(E.77)}$$

The (1,1) and (1, 2) blocks of the Jacobian matrix can be obtained from (E.76) by taking increments of the form $(\delta a, 0)$, $(0, \delta b)$, and the (2,1) and (2, 2) with increments of the type $(\delta a, 0)$, $(0, \delta b)$ in (E.77), respectively.

Remark E.4 In the non-targeted version of the scalar DCC model we consider the parameters $\boldsymbol{\theta} := (a, b, \boldsymbol{c}) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^N$, $N := \frac{1}{2}n(n+1)$ subjected to the positivity an stationarity constraints (2.31)-(2.32), which amount to three linear conditions, namely, (E.67)-(E.69), together with the positive semidefiniteness constraint

$$M(\boldsymbol{c}) := \text{math}(\boldsymbol{c}) \succ 0 \in \mathbb{S}_n. \quad \text{(E.78)}$$

The local penalized model for the optimization problem (3.7) in the scalar non-targeted DCC model case is obtained from (E.70) by adding the term associated to the constraint (E.78), namely, $L_2 D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$,

with L_2 the corresponding penalization strength. The log-likelihood function, its gradient, and the Jacobian of the gradient of the local model are obtained directly from (D.1)-(D.5), and (D.9) in Proposition D.1 taking into account Remarks D.2 and D.4. In relation (E.70) the Bregman divergence associated to the additional constraint (E.78) is obtained by using (3.5):

$$D_M(\boldsymbol{c}, \boldsymbol{c}^{(k)}) = \operatorname{tr}(M(\boldsymbol{c}) \cdot M(\boldsymbol{c}^{(k)})^{-1}) - \log \det(M(\boldsymbol{c}) \cdot M(\boldsymbol{c}^{(k)})^{-1}) - n. \tag{E.79}$$

The gradient of the local model (E.70) with (E.79) has the following third component in the non-targeted case:

$$\nabla_{\boldsymbol{c}} \tilde{f}^{(k)}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{c}} f(\boldsymbol{\theta}^{(k)}) + H_{\boldsymbol{c}}^{(k)}(\boldsymbol{c} - \boldsymbol{c}^{(k)}) - L_2 \ \mathrm{math}^*(\mathrm{math}(\boldsymbol{c})^{-1} - \mathrm{math}(\boldsymbol{c}^{(k)})^{-1}). \tag{E.80}$$

The same applies to the Jacobian of the gradient of the local model. The expressions (E.76) and (E.77) remain valid for any $\boldsymbol{\delta\theta} := (\delta a, \delta b, \delta \boldsymbol{c}) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^N$ and, additionally, we provide

$$T_{\boldsymbol{\theta}} \nabla_{\boldsymbol{c}} \tilde{f}^{(k)}(\boldsymbol{\delta\theta}) = H_{\boldsymbol{c}}^{(k)} \boldsymbol{\delta c} + L_2 \ \mathrm{math}^* \left(\mathrm{math}(\boldsymbol{c})^{-1} \ \mathrm{math}(\boldsymbol{\delta c}) \ \mathrm{math}(\boldsymbol{c})^{-1}\right). \tag{E.81}$$

The relevant blocks of the Jacobian matrix of dimension three are obtained from (E.76), (E.77), and (E.81) by taking the corresponding increments $\boldsymbol{\delta\theta}$. For example, the elements $(1,1)$, $(1,2)$, and $(1,3)$ of the Jacobian require taking increments of the form $(\delta a, 0, 0)$, $(0, \delta b, 0)$, and $(0, 0, \delta c)$, respectively.

F Detailed computations of the results in Section E

F.1 The Hadamard DCC model

The gradient of the local model The computation of $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ requires the expressions of the gradients of each of the divergences in (E.10)-(E.12). We start with the Bregman divergences $D_{M_1}(\boldsymbol{a}, \boldsymbol{a}^{(k)})$, $D_{M_2}(\boldsymbol{b}, \boldsymbol{b}^{(k)}) \in \mathbb{S}_n$ in (E.10)-(E.11) related to the positive semidefiniteness constraints (E.6) and (E.7), respectively. We have

$$\nabla_{\boldsymbol{a}} D_{M_1}(\boldsymbol{a}, \boldsymbol{a}^{(k)}) = -\mathrm{math}^*(M_1(\boldsymbol{a})^{-1} - M_1(\boldsymbol{a}^{(k)})^{-1}), \tag{F.1}$$

$$\nabla_{\boldsymbol{b}} D_{M_2}(\boldsymbol{b}, \boldsymbol{b}^{(k)}) = -\mathrm{math}^*(M_2(\boldsymbol{b})^{-1} - M_2(\boldsymbol{b}^{(k)})^{-1}), \tag{F.2}$$

where $\mathrm{math}^* : \mathbb{S}_n \longrightarrow \mathbb{R}^N$ is the adjoint map of the operator $\mathrm{math} : \mathbb{R}^N \longrightarrow \mathbb{S}_n$, both introduced in Subsection A.2. We now provide the expression of the gradient of the divergence $D_{M_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \in \mathbb{S}_n$ defined in (E.12) and related to the positive definiteness constraint (E.8). By relation (E.2) it can be written down as

$$\nabla_{\boldsymbol{\theta}} D_{M_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = (\mathrm{math}^*((M_3(\boldsymbol{\theta})^{-1} - M_3(\boldsymbol{\theta}^{(k)})^{-1}) \odot S), \mathrm{math}^*((M_3(\boldsymbol{\theta})^{-1} - M_3(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)), \tag{F.3}$$

where we just used that the tangent map $T_{\boldsymbol{\theta}} M_3 : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{S}_n$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^N \times \mathbb{R}^N$ is given by

$$T_{\boldsymbol{\theta}} M_3 \cdot \boldsymbol{\delta\theta} = -\mathrm{math}(\boldsymbol{\delta a} + \boldsymbol{\delta b}) \odot S. \tag{F.4}$$

The substitution of (F.1)-(F.3) into (E.1) gives the components (E.13)-(E.14) of the gradient of the local model (E.9) for the Hadamard family, as required.

The Jacobian of the gradient of the local model The only new element required in the expression (E.5) are the adjoints $T^*_{\boldsymbol{\theta}}M_j : \mathbb{S}_n \longrightarrow \mathbb{R}^N \times \mathbb{R}^N$ of the tangent maps $T_{\boldsymbol{\theta}}M_j : \mathbb{R}^N \times \mathbb{R}^N \longrightarrow \mathbb{S}_n$, $j = 1, 2, 3$. In the case of $j = 1$ and $j = 2$, these adjoints are given by the operator math*. We now compute the expression for the adjoint map $T^*_{\boldsymbol{\theta}}M_3 : \mathbb{S}_n \longrightarrow \mathbb{R}^N \times \mathbb{R}^N$ by dualizing (F.4) for $\Delta \in \mathbb{S}_n$ arbitrary and obtain

$$\begin{aligned}\langle T^*_{\boldsymbol{\theta}}M_3 \cdot \Delta, (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle &= \langle \Delta, T_{\boldsymbol{\theta}}M_3\,(\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = -\mathrm{tr}\,(\Delta((\mathrm{math}(\boldsymbol{\delta a} + \boldsymbol{\delta b}))\odot S)) = -\mathrm{tr}\,((\mathrm{math}(\boldsymbol{\delta a})\odot S)\Delta)\\ &\quad -\mathrm{tr}\,((\mathrm{math}(\boldsymbol{\delta b})\odot S)\Delta) = -\mathrm{tr}\,((\Delta\odot S)\,\mathrm{math}(\boldsymbol{\delta a})) - \mathrm{tr}\,((\Delta\odot S)\,\mathrm{math}(\boldsymbol{\delta b}))\\ &= -\langle \Delta\odot S, \mathrm{math}(\boldsymbol{\delta a})\rangle - \langle \Delta\odot S, \mathrm{math}(\boldsymbol{\delta b})\rangle = -\langle \mathrm{math}^*(\Delta\odot S), \boldsymbol{\delta a}\rangle\\ &\quad -\langle \mathrm{math}^*(\Delta\odot S), \boldsymbol{\delta b}\rangle = \langle (\mathrm{math}^*(\Delta\odot S), \mathrm{math}^*(\Delta\odot S)), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle,\end{aligned}$$

where we used the Hadamard product trace property (A.4). These equalities show that

$$T^*_{\boldsymbol{\theta}}M_3 \cdot \Delta = -\left(\mathrm{math}^*(\Delta\odot S), \mathrm{math}^*(\Delta\odot S)\right). \tag{F.5}$$

The relations (E.15)-(E.16) hence follow from (E.5), as required.

F.2 The rank deficient DCC model

Gradient of the local model The computation of the gradient $\nabla\tilde{f}(\boldsymbol{\theta})$ requires the expressions for the gradients of the components of each of the Bregman divergences defined in (E.29)-(E.31). Regarding the divergences (E.29)-(E.30) associated with the identification constraints we have

$$\nabla_{\boldsymbol{\theta}} D^j_{L_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -\left[\frac{1}{a_{i_j}} - \frac{1}{a^{(k)}_{i_j}}\right]\mathbf{e}^{(i_j)}_{N^*}, \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1, \ldots, r\}, \tag{F.6}$$

$$\nabla_{\boldsymbol{\theta}} D^j_{L_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -\left[\frac{1}{b_{i_j}} - \frac{1}{b^{(k)}_{i_j}}\right]\mathbf{e}^{(i_j)}_{N^*}, \quad i_j = n(j-1) + \frac{1}{2}j(3-j), \quad j \in \{1, \ldots, r\}, \tag{F.7}$$

where we recall that $\mathbf{e}^{(i_j)}_{N^*} \in \mathbb{R}^{N^*}$ is the canonical unit vector whose i_j-th entry equals one. We now proceed with the positive definiteness constraint (E.27), for which we recall that

$$M\,(\boldsymbol{\theta}) = (\mathbf{i}_n\mathbf{i}_n^\top - \mathrm{mat}_r\,(\boldsymbol{a})\,\mathrm{mat}_r\,(\boldsymbol{a})^\top - \mathrm{mat}_r\,(\boldsymbol{b})\,\mathrm{mat}_r\,(\boldsymbol{b})^\top)\odot S.$$

In order to provide the gradient of the Bregman divergence associated to the constraint (E.27), we refer to the relation (E.2), whose explicit formulation requires the expression of the adjoint map $T^*_{\boldsymbol{\theta}}M : \mathbb{S}_n \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$. We hence first determine the corresponding tangent map $T_{\boldsymbol{\theta}}M : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{S}_n$. Consider an arbitrary element $\boldsymbol{\delta\theta} = (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$. Then,

$$T_{\boldsymbol{\theta}}M \cdot \boldsymbol{\delta\theta} = -(\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top + \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top)\odot S.$$

We dualize this expression in order to obtain the required adjoint map. For any element $\Delta \in \mathbb{S}_n$ we have:

$$\begin{aligned}&\langle T^*_{\boldsymbol{\theta}}M \cdot \Delta, (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = \langle \Delta, T_{\boldsymbol{\theta}}M \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle\\ &= -\langle \Delta, (\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top + \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top)\odot S\rangle\\ &= -\langle \Delta, (\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top)\odot S\rangle - \langle \Delta, (\mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top)\odot S\rangle - \langle \Delta, (\mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top)\odot S\rangle\\ &\quad - \langle \Delta, (\mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top)\odot S\rangle = -\langle (\Delta\odot S)\mathrm{mat}_r(\boldsymbol{a}), \mathrm{mat}_r(\boldsymbol{\delta a})\rangle - \langle \mathrm{mat}_r(\boldsymbol{a})^\top(\Delta\odot S), \mathrm{mat}_r(\boldsymbol{\delta a})^\top\rangle\\ &\quad - \langle (\Delta\odot S)\mathrm{mat}_r(\boldsymbol{b}), \mathrm{mat}_r(\boldsymbol{\delta b})\rangle - \langle \mathrm{mat}_r(\boldsymbol{b})^\top(\Delta\odot S), \mathrm{mat}_r(\boldsymbol{\delta b})^\top\rangle = -2\,\langle \mathrm{mat}^*_r((\Delta\odot S)\mathrm{mat}_r(\boldsymbol{a})), \boldsymbol{\delta a}\rangle\\ &\quad - 2\,\langle \mathrm{mat}^*_r((\Delta\odot S)\mathrm{mat}_r(\boldsymbol{b})), \boldsymbol{\delta b}\rangle = -2\langle (\mathrm{mat}^*_r((\Delta\odot S)\mathrm{mat}_r(\boldsymbol{a})), \mathrm{mat}^*_r((\Delta\odot S)\mathrm{mat}_r(\boldsymbol{b}))), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle.\end{aligned}$$

Recall that since by Proposition A.1 mat_r^* equals vec_r, we hence have

$$T_{\boldsymbol{\theta}}^* M \cdot \Delta = -2\ (\mathrm{vec}_r((\Delta \odot S)\mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r((\Delta \odot S)\mathrm{mat}_r(\boldsymbol{b}))) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}. \tag{F.8}$$

If we substitute (F.8) into (E.2) we obtain the following expression for the gradient of the Bregman divergence $D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$:

$$\begin{aligned}\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -T_{\boldsymbol{\theta}}^* M(M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) = 2\ (\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\ \mathrm{mat}_r(\boldsymbol{a})),\\ \mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\ \mathrm{mat}_r(\boldsymbol{b}))).\end{aligned} \tag{F.9}$$

Finally, the relations (F.6), (F.7), and (F.9) lead to the expressions (E.32)-(E.33) for the components of the gradient for the local model (E.28) in the case of the rank deficient family.

Jacobian of the gradient of the local model We use the general expression (E.5) to determine the tangent map of $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ for any $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b})$. The only ingredient that needs to be specified in order to apply (E.5) is the tangent map of the gradient $\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ (F.9) of the Bregman divergence (E.31) associated to the positive definiteness constraint (E.27). A straightforward computation yields

$$\begin{aligned}&T_{\boldsymbol{\theta}}(\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})) \cdot \boldsymbol{\delta\theta} = -T_{\boldsymbol{\theta}}(T_{\boldsymbol{\theta}}^* M(M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1})) \cdot \boldsymbol{\delta\theta}\\ &= 2\ T_{\boldsymbol{\theta}}\left(\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{b}))\right) \cdot \boldsymbol{\delta\theta}\\ &= -2(\mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1}(T_{\boldsymbol{\theta}} M \cdot \boldsymbol{\delta\theta}) M(\boldsymbol{\theta})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{a}) - ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{\delta a})),\\ &\qquad \mathrm{vec}_r(((M(\boldsymbol{\theta})^{-1}(T_{\boldsymbol{\theta}} M \cdot \boldsymbol{\delta\theta}) M(\boldsymbol{\theta})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{b}) - ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\mathrm{mat}_r(\boldsymbol{\delta b}))).\end{aligned} \tag{F.10}$$

The substitution into (E.5) of the expressions for the tangent maps and their adjoints provided in this appendix yields the relations (E.34)-(E.35) that can be used to obtain the blocks of the Jacobian matrix for the local model (E.28) in the case of the rank deficient DCC specification.

Remark F.1 When working with the non-targeted version of this model, we need to add the Bregman divergence (E.39) to make the extra parameter positive definite. Its gradient is obtained analogously to (F.1)-(F.2). We now carry out the computation of the gradients of the Bregman divergences (E.40) related to the stationarity constraints (E.37)-(E.38) that need to be added separately in this case by using the relation (E.3). In order to provide explicit expressions for those gradients, we first compute the tangent maps $T_{\boldsymbol{\theta}} N_j : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{R}^N$, $j \in \{1, 2\}$:

$$\begin{aligned}T_{\boldsymbol{\theta}} N_j \cdot \boldsymbol{\delta\theta} =& (-1)^j \cdot \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top\\ &+ \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top),\end{aligned} \tag{F.11}$$

with $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}, \boldsymbol{\delta c}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \times \mathbb{R}^N$. We determine their adjoints $T_{\boldsymbol{\theta}}^* N_j : \mathbb{R}^N \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \times \mathbb{R}^N$, $j \in \{1, 2\}$ by dualizing (F.11). Consider an arbitrary vector $\mathbf{v} \in \mathbb{R}^N$. We then have

$$\begin{aligned}&\langle T_{\boldsymbol{\theta}}^* N_j \cdot \mathbf{v}, \boldsymbol{\delta\theta}\rangle = \langle \mathbf{v}, T_{\boldsymbol{\theta}} N_j\ (\boldsymbol{\delta\theta})\rangle = (-1)^j \langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top + \mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top\\ &+ \mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top + \mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top)\rangle = (-1)^j(\langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{\delta a})\mathrm{mat}_r(\boldsymbol{a})^\top)\rangle\\ &+ \langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{a})\mathrm{mat}_r(\boldsymbol{\delta a})^\top)\rangle + \langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{\delta b})\mathrm{mat}_r(\boldsymbol{b})^\top)\rangle + \langle \mathbf{v}, \mathrm{vech}(\mathrm{mat}_r(\boldsymbol{b})\mathrm{mat}_r(\boldsymbol{\delta b})^\top)\rangle)\\ &= (-1)^j(\langle \mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{a}), \mathrm{mat}_r(\boldsymbol{\delta a})\rangle + \langle \mathrm{mat}_r(\boldsymbol{a})^\top \mathrm{vech}^*(\mathbf{v}), \mathrm{mat}_r(\boldsymbol{\delta a})^\top\rangle + \langle \mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{b}), \mathrm{mat}_r(\boldsymbol{\delta b})\rangle\\ &+ \langle \mathrm{mat}_r(\boldsymbol{b})^\top \mathrm{vech}^*(\mathbf{v}), \mathrm{mat}_r(\boldsymbol{\delta b})^\top\rangle) = 2 \cdot (-1)^j(\langle \mathrm{mat}_r^*(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{a})), \boldsymbol{\delta a}\rangle\\ &+ \langle \mathrm{mat}_r^*(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{b})), \boldsymbol{\delta b}\rangle) = 2 \cdot (-1)^j \langle (\mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{b})), (\boldsymbol{\delta a}, \boldsymbol{\delta b}))\rangle,\end{aligned}$$

where we used that, by Proposition A.1, mat_r^* equals vec_r. We hence obtain that

$$T^*_{\boldsymbol{\theta}} N_j \cdot \mathbf{v} = 2 \cdot (-1)^j (\mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{a})), \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{v})\mathrm{mat}_r(\boldsymbol{b})), \quad j \in \{1,2\}\,. \tag{F.12}$$

Consequently, the gradients of the components of the Bregman divergences $D^i_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, $i \in \{1,\dots,N\}$, $j \in \{1,2\}$ are given by the relations

$$\begin{aligned}\nabla_{\boldsymbol{\theta}} D^i_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) =& 2 \cdot (-1)^{j-1} \left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i} \right] \\ & \times (\mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}^{(i)}_N) \cdot \mathrm{mat}_r(\boldsymbol{a}), \mathrm{vec}_r(\mathrm{vech}^*(\mathbf{e}^{(i)}_N) \cdot \mathrm{mat}_r(\boldsymbol{b})),\end{aligned} \tag{F.13}$$

where $\mathbf{e}^{(i)}_N \in \mathbb{R}^N$, $i \in \{1,\dots,N\}$ are the canonical unit vectors.

F.3 The Almon DCC model

Gradient of the local model We determine the explicit expression for the gradient of the local model which requires the computation of the gradients of each of the Bregman divergences presented in (E.49)-(E.51). Concerning the divergences associated with the nonlinear positivity constraints we use the relation (E.3) and for the divergences $D_{N_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ and $D_{N_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ related to the inequalities (E.45)-(E.46) we write

$$\nabla_{\boldsymbol{\theta}} D_{N_1}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = - \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{a}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{a}^{(k)}))_1} \right] \left(K_{\boldsymbol{a}}^\top \cdot \mathbf{e}^{(1)}_n \right), \tag{F.14}$$

$$\nabla_{\boldsymbol{\theta}} D_{N_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = - \left[\frac{1}{(\mathrm{alm}_n(\boldsymbol{b}))_1} - \frac{1}{(\mathrm{alm}_n(\boldsymbol{b}^{(k)}))_1} \right] \left(K_{\boldsymbol{b}}^\top \cdot \mathbf{e}^{(1)}_n \right), \tag{F.15}$$

where we used the expression (A.17) of the adjoint $T^*_{\mathbf{v}}\mathrm{alm}_n : \mathbb{R}^n \longrightarrow \mathbb{R}^3$ of the tangent map $T_{\mathbf{v}}\mathrm{alm}_n : \mathbb{R}^3 \longrightarrow \mathbb{R}^n$; recall that $K_{\boldsymbol{a}}, K_{\boldsymbol{b}} \in \mathbb{M}_{n,3}$ are the matrices introduced in (A.16) and $\mathbf{e}^{(1)}_n \in \mathbb{R}^n$ is the canonical unit vector. We now proceed with the positive definiteness constraint (E.27), for which we recall that

$$M(\boldsymbol{\theta}) = (\mathbf{i}_n \mathbf{i}_n^\top - \mathrm{alm}_n(\boldsymbol{a})\,\mathrm{alm}_n(\boldsymbol{a})^\top - \mathrm{alm}_n(\boldsymbol{b})\,\mathrm{alm}_n(\boldsymbol{b})^\top) \odot S.$$

In order to compute the gradient of the Bregman divergence associated to the constraint (E.27), we refer to the relation (E.2), whose explicit formulation requires the expression of the adjoint map $T^*_{\boldsymbol{\theta}} M : \mathbb{S}_n \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$. We hence first determine the corresponding tangent map $T_{\boldsymbol{\theta}} M : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \longrightarrow \mathbb{S}_n$; consider an arbitrary element $\boldsymbol{\delta\theta} = (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^{N^*} \times \mathbb{R}^{N^*}$, then

$$T_{\boldsymbol{\theta}} M \cdot \boldsymbol{\delta\theta} = -((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top + \mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top + (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^\top + \mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top) \odot S$$

and dualize it in order to obtain the required adjoint, that is, for arbitrary $\Delta \in \mathbb{S}_n$, we compute

$$\begin{aligned}&\langle T^*_{\boldsymbol{\theta}} M \cdot \Delta, (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle = \langle \Delta, T_{\boldsymbol{\theta}} M \cdot (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle \\ &= -\langle \Delta, ((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top + \mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top + (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^\top + \mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top) \odot S\rangle \\ &= -\langle \Delta, ((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top) \odot S\rangle - \langle \Delta, (\mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top) \odot S\rangle - \langle \Delta, ((K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^\top) \odot S\rangle \\ &\quad - \langle \Delta, (\mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top) \odot S\rangle = -\langle (\Delta \odot S)\mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}\rangle - \langle \mathrm{alm}_n(\boldsymbol{a})^\top (\Delta \odot S), (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top\rangle \\ &\quad - \langle (\Delta \odot S)\mathrm{alm}_n(\boldsymbol{b}), K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}\rangle - \langle \mathrm{alm}_n(\boldsymbol{b})^\top (\Delta \odot S), (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top\rangle = -2\,\langle K_{\boldsymbol{a}}^\top (\Delta \odot S)\mathrm{alm}_n(\boldsymbol{a}), \boldsymbol{\delta a}\rangle \\ &\quad - 2\,\langle K_{\boldsymbol{b}}^\top (\Delta \odot S)\mathrm{alm}_n(\boldsymbol{b}), \boldsymbol{\delta b}\rangle = -2\langle (K_{\boldsymbol{a}}^\top (\Delta \odot S)\mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top (\Delta \odot S)\mathrm{alm}_n(\boldsymbol{b})), (\boldsymbol{\delta a}, \boldsymbol{\delta b})\rangle.\end{aligned}$$

This relation is equivalent to

$$T^*_{\boldsymbol{\theta}} M \cdot \Delta = -2\ (K_{\boldsymbol{a}}^\top(\Delta \odot S)\mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top(\Delta \odot S)\mathrm{alm}_n(\boldsymbol{b})) \in \mathbb{R}^3 \times \mathbb{R}^3. \tag{F.16}$$

Substituting directly (F.16) into (E.2) yields the following expression for the gradient of the Bregman divergence $D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$

$$\begin{aligned}\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = -T^*_{\boldsymbol{\theta}} M(M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) = 2\ (&K_{\boldsymbol{a}}^\top((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\ \mathrm{alm}_n(\boldsymbol{a}),\\ &K_{\boldsymbol{b}}^\top((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\ \mathrm{alm}_n(\boldsymbol{b})).\end{aligned} \tag{F.17}$$

Finally, the relations (F.14), (F.15), and (F.17) lead to the expressions (E.52)-(E.53) for the components of the gradient of the local model (E.48) in the case of the Almon family of DCC models.

Jacobian of the gradient of the local model We use the general expression (E.5) to determine the tangent map of $\nabla_{\boldsymbol{\theta}} \tilde{f}^{(k)}(\boldsymbol{\theta})$ for any increment $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}) \in \mathbb{R}^3 \times \mathbb{R}^3$. In order to apply (E.5), we only need to specify the tangent map of the gradient $\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ (F.17) of the Bregman divergence (E.51) associated to the positive definite constraint (E.47). A straightforward computation yields

$$\begin{aligned}
&T_{\boldsymbol{\theta}}(\nabla_{\boldsymbol{\theta}} D_M(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})) \cdot \boldsymbol{\delta\theta} = -T_{\boldsymbol{\theta}}(T^*_{\boldsymbol{\theta}} M(M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1})) \cdot \boldsymbol{\delta\theta}\\
&= 2\ T_{\boldsymbol{\theta}}(K_{\boldsymbol{a}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{b}))\\
&= 2\big((K_{\boldsymbol{aa}} \cdot \boldsymbol{\delta a})^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{a}) - K_{\boldsymbol{a}} \cdot ((M(\boldsymbol{\theta})^{-1}(T_{\boldsymbol{\theta}} M \cdot \boldsymbol{\delta\theta}) M(\boldsymbol{\theta})^{-1}) \odot S)\\
&\quad \cdot \mathrm{alm}_n(\boldsymbol{a}) + K_{\boldsymbol{a}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}), (K_{\boldsymbol{bb}} \cdot \boldsymbol{\delta b})^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\\
&\quad \cdot \mathrm{alm}_n(\boldsymbol{b}) + K_{\boldsymbol{b}} \cdot ((-M(\boldsymbol{\theta})^{-1}(T_{\boldsymbol{\theta}} M \cdot \boldsymbol{\delta\theta}) M(\boldsymbol{\theta})^{-1}) \odot S) \cdot \mathrm{alm}_n(\boldsymbol{b}) + K_{\boldsymbol{b}}^\top \cdot ((M(\boldsymbol{\theta})^{-1} - M(\boldsymbol{\theta}^{(k)})^{-1}) \odot S)\\
&\quad \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})\big),
\end{aligned} \tag{F.18}$$

where

$$\begin{aligned}
K_{\boldsymbol{aa}} \cdot \boldsymbol{\delta a} &:= (\mathbf{0}_n \mid \mathbf{k}_n^1 \odot (K_{\boldsymbol{a}}^0 \cdot \boldsymbol{\delta a}) \mid \mathbf{k}_n^2 \odot (K_{\boldsymbol{a}}^0 \cdot \boldsymbol{\delta a})),\\
K_{\boldsymbol{bb}} \cdot \boldsymbol{\delta b} &:= (\mathbf{0}_n \mid \mathbf{k}_n^1 \odot (K_{\boldsymbol{b}}^0 \cdot \boldsymbol{\delta b}) \mid \mathbf{k}_n^2 \odot (K_{\boldsymbol{b}}^0 \cdot \boldsymbol{\delta b})),\\
K_{\boldsymbol{a}}^0 &:= \left(\mathbf{0}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\boldsymbol{a}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\boldsymbol{a}})\right),\\
K_{\boldsymbol{b}}^0 &:= \left(\mathbf{0}_n \mid \mathbf{k}_n^1 \odot \mathrm{alm}_n(\bar{\boldsymbol{b}}) \mid \mathbf{k}_n^2 \odot \mathrm{alm}_n(\bar{\boldsymbol{b}})\right).
\end{aligned}$$

The substitution of the expressions for the tangent maps and their adjoints provided in this appendix into (E.5) yields the relations (E.54)-(E.55) that can be used to obtain the blocks of the Jacobian matrix for the local model (E.48) in the case of the Almon DCC specification.

Remark F.2 In the non-targeted version of the Almon DCC model the parameters $\boldsymbol{\theta} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^N$ are subjected to the identification constraints (E.45), (E.46), to the stationarity constraints (E.56)-(E.57), and to the positivity constraint (E.58). For the first two constraints the computations are given above, for the last one they can be obtained analogously to the ones provided for (E.10)-(E.11) in Section F.1. As for the stationarity constraints, we now carry out the necessary computation. We start by computing the gradients of the components $D^i_{N_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}), D^i_{N_4}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, $i \in \{1, \ldots, N\}$ of the Bregman divergences associated to the nonlinear positivity constraints in (E.56) and (E.57), respectively. In order to do that we first need to compute the maps $T_{\boldsymbol{\theta}} N_j : \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \times \mathbb{R}^N \longrightarrow \mathbb{R}^N$, $j \in \{3, 4\}$. Let $\boldsymbol{\delta\theta} := (\boldsymbol{\delta a}, \boldsymbol{\delta b}, \boldsymbol{\delta c}) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^N$ arbitrary, then the following relations hold true:

$$\begin{aligned}
T_{\boldsymbol{\theta}} N_j \cdot \boldsymbol{\delta\theta} = (-1)^j \cdot \mathrm{vech}((&K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top + \mathrm{alm}_n(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top\\
&+ (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n(\boldsymbol{b})^\top + \mathrm{alm}_n(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top), \quad j \in \{3, 4\}.
\end{aligned} \tag{F.19}$$

In order to use (E.3), we need to determine the adjoint maps $T^*_{\boldsymbol{\theta}}N_j : \mathbb{R}^N \longrightarrow \mathbb{R}^{N^*} \times \mathbb{R}^{N^*} \times \mathbb{R}^N$ by dualizing the relations (F.19). Let $\mathbf{v} \in \mathbb{R}^N$ arbitrary; we have

$$\begin{aligned}
&\langle T^*_{\boldsymbol{\theta}}N_j \cdot \mathbf{v}, \boldsymbol{\delta\theta}\rangle = \langle \mathbf{v}, T_{\boldsymbol{\theta}}N_j\,(\boldsymbol{\delta\theta})\rangle = (-1)^j\langle \mathbf{v}, \mathrm{vech}((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top + \mathrm{alm}_n\,(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top \\
&\quad + (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n\,(\boldsymbol{b})^\top + \mathrm{alm}_n\,(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top)\rangle = (-1)^j(\langle \mathbf{v}, \mathrm{vech}((K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}) \cdot \mathrm{alm}_n(\boldsymbol{a})^\top)\rangle \\
&\quad + \langle \mathbf{v}, \mathrm{vech}(\mathrm{alm}_n\,(\boldsymbol{a}) \cdot (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top)\rangle + \langle \mathbf{v}, \mathrm{vech}((K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}) \cdot \mathrm{alm}_n\,(\boldsymbol{b})^\top)\rangle + \langle \mathbf{v}, \mathrm{vech}(\mathrm{alm}_n\,(\boldsymbol{b}) \cdot (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top)\rangle) \\
&=(-1)^j\Big(\langle \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a}\rangle + \langle \mathrm{alm}_n(\boldsymbol{a})^\top \cdot \mathrm{vech}^*(\mathbf{v}), (K_{\boldsymbol{a}} \cdot \boldsymbol{\delta a})^\top\rangle + \langle \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{b}), K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b}\rangle \\
&\quad + \langle \mathrm{alm}_n(\boldsymbol{b})^\top \cdot \mathrm{vech}^*(\mathbf{v}), (K_{\boldsymbol{b}} \cdot \boldsymbol{\delta b})^\top\rangle\Big) = 2(-1)^j\Big(\langle K_{\boldsymbol{a}}^\top \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{a}), \boldsymbol{\delta a}\rangle + \langle K_{\boldsymbol{b}}^\top \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{b}), \boldsymbol{\delta b}\rangle\Big) \\
&=2(-1)^j\langle (K_{\boldsymbol{a}}^\top \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{b})), \boldsymbol{\delta\theta}\rangle, \quad j \in \{3,4\},
\end{aligned}$$

which immediately yields that

$$T^*_{\boldsymbol{\theta}}N_j \cdot \mathbf{v} = 2(-1)^j\Big(K_{\boldsymbol{a}}^\top \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{a}), K_{\boldsymbol{b}}^\top \cdot \mathrm{vech}^*(\mathbf{v}) \cdot \mathrm{alm}_n(\boldsymbol{b})\Big), \quad j \in \{3,4\}.$$

We hence obtain the following expression for the components of the gradients of the divergences $D_{N_3}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ and $D_{N_4}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$:

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} D^i_{N_j}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = 2 \cdot (-1)^{j-1} \cdot \Bigg(&\left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right] (K_{\boldsymbol{a}}^\top \cdot \mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{alm}_n(\boldsymbol{a})), \\
&\left[\frac{1}{(N_j(\boldsymbol{\theta}))_i} - \frac{1}{(N_j(\boldsymbol{\theta}^{(k)}))_i}\right] (K_{\boldsymbol{b}}^\top \cdot \mathrm{vech}^*(\mathbf{e}_N^{(i)}) \cdot \mathrm{alm}_n(\boldsymbol{b}))\Bigg), \quad j \in \{3,4\}\,.
\end{aligned} \tag{F.20}$$

G Additional empirical results

This section completes the results in Section 4 of the paper and also reports analogous empirical findings for a dataset based on the same assets but on a different time period (called **Period I**). In the case of Period I, the price quotes are taken from January 19th, 1996 to December 21st, 2010 which amounts to 3750 observations in the sample. The first 3000 observations (January 19th, 1996 - December 31st, 2007) are reserved for model estimation and the remaining 750 are used for an out-of-sample study. Notice that Period I contains the 2008-09 crisis events in the out-of-sample interval, while the dataset used in the main body of the paper (referred to as **Period II**) includes these events exclusively in the part of the sample used for estimation purposes.

Figure 1 complements Figure 2 in the main body of the paper and plots, for Period II, the values of the entries of the parameter vector $\boldsymbol{b}$ for of the rank one deficient and Almon shuffle models in dimension 30 with the assets ordered according to the Almon shuffle estimation. The figure also shows the corresponding estimate implied by the scalar model, which is the square root of the parameter b in (2.30), and it illustrates the flexibility of the rank one deficient and Almon models with respect to the scalar one. The pattern of estimates of the entries of $\boldsymbol{b}$ of the Almon model is convex.

The entries of $\boldsymbol{a}$ and $\boldsymbol{b}$ in the Almon shuffle model for both Figure 2 in the paper and Figure 1 are computed by using the Almon function defined in Section 2.4 using the estimates of the three intrinsic parameters of Almon function for each of the dimensions considered. Table G.1 reports on these Almon parameters for Period I and Period II.

Tables G.2 up to G.10 correspond to Period I and are analogous to the tables provided in Section 4 of the paper for Period II. More specifically, Tables G.2, G.3, G.4, G.5-G.6, G.7-G.8, and G.9-G.10 (for

DJIA dataset. Period I

n	a_1	a_2	a_3	b_1	b_2	b_3
5	-0.97647	-0.02374	0.01028	0.01344	-0.02047	0.00293
10	-0.94870	-0.00176	0.00065	-0.00758	0.00247	-0.00042
15	-0.96304	0.00090	0.00019	-0.01338	0.00346	-0.00030
20	-0.97398	0.00448	-0.00015	-0.00434	-0.00001	-0.00002
25	-0.96169	0.00099	-0.00001	-0.01266	0.00112	-0.00004
30	-0.96930	0.00124	-0.00001	-0.01275	0.00121	-0.00004

DJIA dataset. Period II

n	a_1	a_2	a_3	b_1	b_2	b_3
5	-0.78996	-0.07976	0.00909	-0.02519	0.01330	-0.00232
10	-0.87945	-0.01598	0.00117	-0.00609	-0.00197	0.00014
15	-0.90004	-0.01347	0.00090	-0.00926	0.00122	-0.00012
20	-0.93593	0.00045	-0.00007	-0.00144	-0.00152	0.00007
25	-0.94422	0.00077	-0.00005	-0.00140	-0.00112	0.00004
30	-0.95261	0.00113	-0.00004	-0.00088	-0.00095	0.00003

Table G.1: Estimates of the parameters of the Almon functions $(\text{alm}_n(\boldsymbol{a}))_i = a_1 + \exp(a_2 i + a_3 i^2)$ and $(\text{alm}_n(\boldsymbol{b}))_i = b_1 + \exp(b_2 i + b_3 i^2)$ for the Almon shuffle models. The symbol i represents the entry number and n is the dimension of the model.

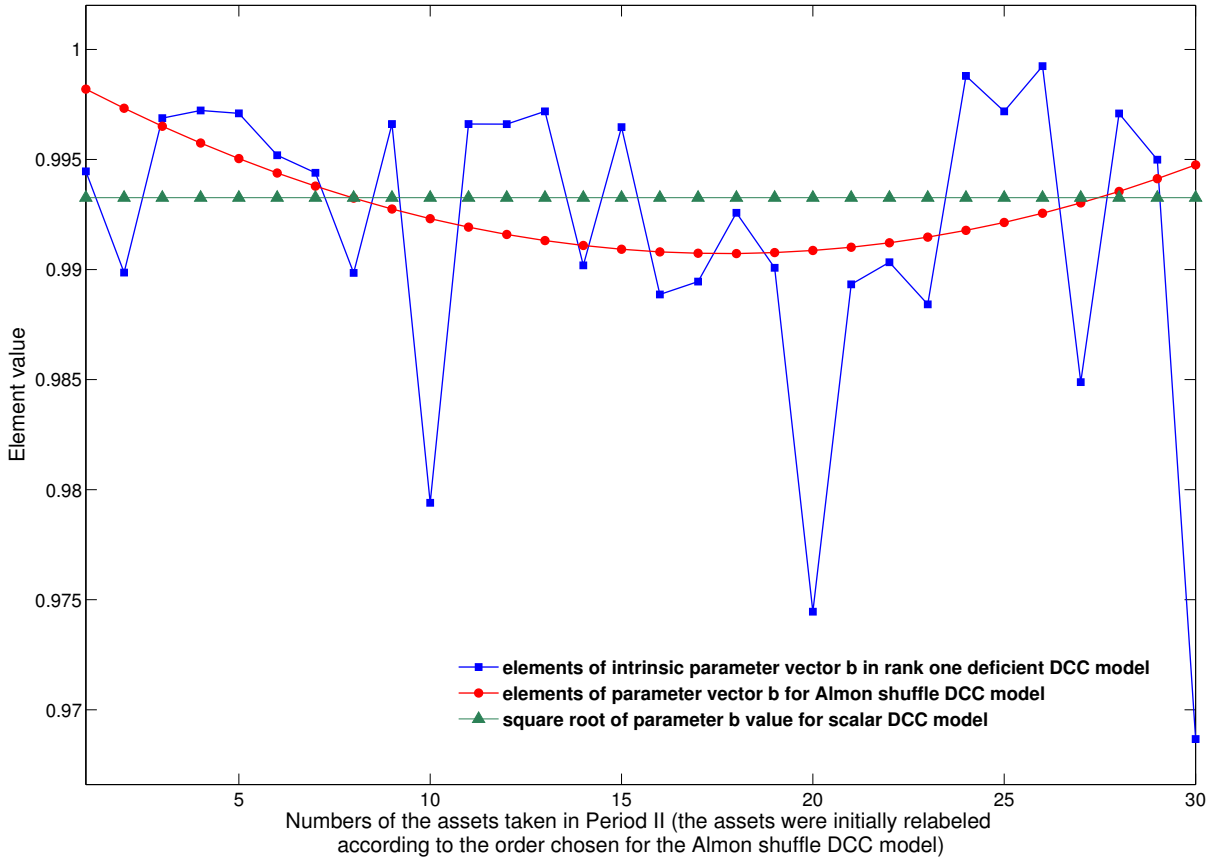

Figure 1: Estimates of the entries of the parameter $\boldsymbol{b}$ of the rank one deficient and Almon shuffle models for the thirty assets case in the order determined by the Almon shuffle estimation (dataset of Period II).

Period I) correspond to Tables 4.2, 4.3, 4.5, 4.6-4.7, 4.8-4.9, 4.10-4.11 (for Period II), respectively.

DJIA dataset. Period I.

n		Scalar	Almon	Almon Shuffle	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Hadamard
5	-log$L$	-13.9889	-13.9892	-13.9912	-13.9926	-13.9932	**-13.9945**
	AIC$^{\text{rank}}$	-27.9665^{3}	-27.9644^{4}	-27.9683^{2}	**-27.9685**^{1}	-27.9643^{5}	-27.9591^{6}
10	-log$L$	-28.3038	-28.3042	-28.3038	-28.3102	-28.3130	**-28.3163**
	AIC$^{\text{rank}}$	-56.5862^{2}	-56.5843^{3}	-56.5835^{4}	**-56.5870**^{1}	-56.5807^{5}	-56.5392^{6}
15	-log$L$	-42.0486	-42.0489	-42.0510	-42.0592	-42.0670	**-42.0725**
	AIC$^{\text{rank}}$	-84.0659^{3}	-84.0637^{5}	-84.0680^{2}	**-84.0684**^{1}	-84.0653^{4}	-83.9550^{6}
20	-log$L$	-56.6593	-56.6597	-56.6613	-56.6742	-56.6855	**-56.6951**
	AIC$^{\text{rank}}$	-113.2773^{4}	-113.2753^{5}	-113.2786^{3}	**-113.2818**^{1}	-113.2790^{2}	-113.0702^{6}
25	-log$L$	-71.1817	-71.1819	-71.1835	-71.1985	-71.2108	**-71.2296**
	AIC$^{\text{rank}}$	-142.3120^{3}	-142.3098^{4}	-142.3129^{2}	**-142.3137**^{1}	-142.3062^{5}	-141.9760^{6}
30	-log$L$	-85.9232	-85.9253	-85.9267	-85.9523	-85.9599	**-86.0091**
	AIC$^{\text{rank}}$	-171.7851^{4}	-171.7866^{3}	-171.7894^{2}	**-171.8047**^{1}	-171.7811^{5}	-171.3381^{6}
	AIC score$^{\text{rank}}$	19^{3}	24^{4}	15^{2}	**6**^{1}	26^{5}	36^{6}

Table G.2: Normalized values of minus the log-likelihood function (-log$(L)/T_{est}$) and associated AIC statistics. The smallest values of minus the log-likelihood function are displayed in black bold. Exponents on the AIC row indicate the rank of the model from 6 (the worse) to 1 (the best). The "AIC score" row at the bottom contains aggregated ranks by models. Figures in red point to the model that exhibits the lowest AIC value.

DJIA dataset. Period I. Matrix A

n		MacGyver	Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	mean (median)	0.0111 (0.0077)	0.0040	0.0083	0.0030	0.0042	0.0048	0.0046
	std	0.0120	-	0.0078	0.0060	0.0071	0.0015	0.0069
	min	0.0000	0.0040	-0.0012	-0.0018	-0.0014	0.0018	0.0001
	max	0.0559	0.0040	0.0243	0.0172	0.0216	0.0070	0.0296
10	mean (median)	0.0097 (0.0069)	0.0046	0.0056	0.0076	0.0044	0.0050	0.0045
	std	0.0102	-	0.0046	0.0055	0.0037	0.0011	0.0016
	min	0.0000	0.0046	-0.0021	0.0024	-0.0015	0.0035	0.0025
	max	0.0466	0.0046	0.0279	0.0437	0.0193	0.0083	0.0101
15	mean (median)	0.0095 (0.0064)	0.0036	0.0038	0.0039	0.0032	0.0036	0.0036
	std	0.0105	-	0.0038	0.0026	0.0025	0.0004	0.0015
	min	0.0000	0.0036	-0.0030	0.0005	-0.0034	0.0024	0.0014
	max	0.0559	0.0036	0.0237	0.0172	0.0146	0.0042	0.0088
20	mean (median)	0.0088 (0.0059)	0.0026	0.0025	0.0029	0.0025	0.0026	0.0026
	std	0.0100	-	0.0024	0.0015	0.0020	0.0001	0.0006
	min	0.0000	0.0026	-0.0025	0.0006	-0.0034	0.0023	0.0009
	max	0.0559	0.0026	0.0128	0.0088	0.0100	0.0029	0.0035
25	mean (median)	0.0092 (0.0061)	0.0024	0.0023	0.0033	0.0029	0.0023	0.0024
	std	0.0099	-	0.0023	0.0020	0.0021	0.0000	0.0004
	min	0.0000	0.0024	-0.0037	0.0004	-0.0036	0.0023	0.0015
	max	0.0559	0.0024	0.0139	0.0162	0.0141	0.0024	0.0033
30	mean (median)	0.0091 (0.0060)	0.0021	0.0019	0.0028	0.0024	0.0021	0.0022
	std	0.0100	-	0.0023	0.0018	0.0012	0.0001	0.0006
	min	0.0000	0.0021	-0.0047	-0.0003	0.0005	0.0018	0.0010
	max	0.0578	0.0021	0.0127	0.0100	0.0090	0.0023	0.0036

DJIA dataset. Period I. Matrix B

n		MacGyver	Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	mean (median)	0.7731 (0.9484)	0.9764	0.9354	0.9873	0.9774	0.9692	0.9696
	std	0.3311	-	0.0439	0.0115	0.0106	0.0107	0.0088
	min	0.0000	0.9764	0.8700	0.9576	0.9560	0.9539	0.9584
	max	0.9974	0.9764	0.9803	0.9994	0.9990	0.9943	0.9921
10	mean (median)	0.8623 (0.9698)	0.9801	0.9745	0.8791	0.9804	0.9769	0.9797
	std	0.2744	-	0.0081	0.1978	0.0094	0.0069	0.0097
	min	0.0001	0.9801	0.9580	0.2466	0.9600	0.9606	0.9509
	max	0.9974	0.9801	0.9917	0.9955	0.9971	0.9877	0.9921
15	mean (median)	0.8707 (0.9652)	0.9805	0.9781	0.9701	0.9810	0.9799	0.9793
	std	0.2372	-	0.0069	0.0230	0.0130	0.0014	0.0106
	min	0.0000	0.9805	0.9607	0.8915	0.9305	0.9777	0.9434
	max	0.9974	0.9805	0.9935	0.9972	0.9965	0.9835	0.9933
20	mean (median)	0.8514 (0.9656)	0.9871	0.9862	0.9756	0.9854	0.9865	0.9865
	std	0.2717	-	0.0030	0.0247	0.0080	0.0012	0.0029
	min	0.0000	0.9871	0.9795	0.8382	0.9580	0.9847	0.9780
	max	0.9977	0.9871	0.9935	0.9975	0.9970	0.9907	0.9913
25	mean (median)	0.8398 (0.9610)	0.9857	0.9851	0.9589	0.9748	0.9861	0.9847
	std	0.2741	-	0.0027	0.0436	0.0200	0.0014	0.0028
	min	0.0000	0.9857	0.9792	0.7953	0.8685	0.9833	0.9766
	max	0.9977	0.9857	0.9926	0.9970	0.9970	0.9900	0.9892
30	mean (median)	0.8364 (0.9617)	0.9864	0.9857	0.9689	0.9779	0.9866	0.9841
	std	0.2781	-	0.0022	0.0272	0.0159	0.0026	0.0047
	min	0.0000	0.9864	0.9795	0.8425	0.9033	0.9828	0.9674
	max	0.9977	0.9864	0.9932	0.9971	0.9958	0.9960	0.9911

Table G.3: Period I. Estimated parameter matrices A and B of (2.8). Mean, median, standard deviation, minimum and maximum of the entries of the estimated matrices are reported. "MacGyver" stands for the method with this name introduced in [Eng08] based on the use of bivariate scalar DCC models; in this case the median values of estimates are reported between parentheses.

DJIA dataset. Period I. MCS for the correlation of the standardized returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	**1**	**6**	**3**	**5**	**2**	**4**
	p-value	1.000	0.066	0.170	0.170	0.170	0.170
10	Position	**2**	**6**	**1**	**3**	**4**	**5**
	p-value	0.108	0.085	1.000	0.108	0.108	0.108
15	Position	**2**	6	**1**	**4**	**3**	**5**
	p-value	0.683	0.001	1.000	0.159	0.159	0.159
20	Position	**1**	6	3	5	2	4
	p-value	1.000	0.001	0.025	0.025	0.025	0.025
25	Position	2	6	5	3	**1**	4
	p-value	0.004	0.000	0.004	0.004	1.000	0.004
30	Position	2	6	5	3	**1**	4
	p-value	0.000	0.000	0.000	0.000	1.000	0.000
	Score	10	36	18	23	13	26

Table G.4: Model confidence sets (MCS) constructed using the loss function (4.2) based on the correlation of the standardized returns defined in (4.3). For each model and dimension, the integer value in the first row indicates the order of elimination of the model from the MCS (6 stands for the first eliminated model, 5 for the second eliminated model, and so on). In the second row we report the p-value of the test leading to the decision of eliminating the given model from the MCS. The set of integer values printed in bold red identifies the MCS at the confidence level of 90%. The union of integer values printed in bold black and bold red identifies the MCS at the confidence level of 95%. The score of each model in the last row is the sum of the integer values of the six dimensions.

DJIA dataset. Period I. Engle-Colacito regression test for the minimum variance portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	t-stats	**3.76***	4.04*	4.13*	4.09*	3.81*	3.98*
	p-value	0.00	0.00	0.00	0.00	0.00	0.00
10	t-stats	1.73	1.80	1.67	1.90	1.74	**1.65**
	p-value	0.08	0.07	0.10	0.06	0.08	0.10
15	t-stats	-0.55	-0.16	**0.05**	0.26	-0.45	-0.16
	p-value	0.58	0.87	0.96	0.80	0.65	0.87
20	t-stats	-0.51	**-0.02**	0.44	0.43	-0.43	-0.22
	p-value	0.61	0.99	0.66	0.67	0.66	0.82
25	t-stats	**1.47**	2.17*	2.48*	2.15*	1.56	1.96
	p-value	0.14	0.03	0.01	0.03	0.12	0.05
30	t-stats	**2.70***	3.73*	3.85*	3.60*	2.97*	3.43*
	p-value	0.01	0.00	0.00	0.00	0.00	0.00

Table G.5: Results of the Engle-Colacito regression test. The t-stat values refer to the intercept λ of the Engle-Colacito regression (4.4) obtained using a HAC estimator. The p-value can be used to test the null hypothesis that $\lambda = 0$. The symbol * (respectively **) indicates rejection at the 5% (respectively 1%) significance level. Values corresponding to models that exhibit the maximum p-value for a given portfolio cardinality are printed in red.

DJIA dataset. Period I. Engle-Colacito regression test for the equally weighted portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	t-stats	**2.72***	2.80*	2.87*	2.78*	2.77*	2.82*
	p-value	0.01	0.01	0.00	0.01	0.01	0.00
10	t-stats	1.99*	1.91	**1.84**	2.01*	2.06*	1.91
	p-value	0.05	0.06	0.07	0.04	0.04	0.06
15	t-stats	**-1.21**	-1.57	-1.24	-1.44	-1.22	-1.29
	p-value	0.23	0.12	0.22	0.15	0.22	0.20
20	t-stats	-2.32*	-2.89*	**-2.16***	-2.84*	-2.29*	-2.28*
	p-value	0.02	0.00	0.03	0.00	0.02	0.02
25	t-stats	-1.94	-2.39*	**-1.68**	-2.36*	-1.86	-1.76
	p-value	0.05	0.02	0.09	0.02	0.06	0.08
30	t-stats	-3.45*	-3.88*	-3.10*	**-3.01***	-3.18*	-3.12*
	p-value	0.00	0.00	0.00	0.00	0.00	0.00

Table G.6: Results of the Engle-Colacito regression test. The t-stat values refer to the intercept λ of the Engle-Colacito regression (4.4) obtained using a HAC estimator. The p-value can be used to test the null hypothesis that $\lambda = 0$. The symbol * (respectively **) indicates rejection at the 5% (respectively 1%) significance level. Values corresponding to models that exhibit the maximum p-value for a given portfolio cardinality are printed in red.

DJIA dataset. Period I. MCS of EPA for the minimum variance portfolio squared returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	**3**	**2**	**4**	**1**	**5**	**6**
	p-value	0.856	0.856	0.810	1.000	0.810	0.810
10	Position	**4**	**2**	**1**	**3**	**5**	**6**
	p-value	0.335	0.579	1.000	0.335	0.335	0.270
15	Position	**5**	**4**	**1**	**3**	**6**	**2**
	p-value	0.374	0.588	1.000	0.600	0.290	0.600
20	Position	**2**	**5**	**4**	**6**	**3**	**1**
	p-value	0.206	0.206	0.206	0.206	0.206	1.000
25	Position	**3**	**6**	**2**	**5**	**4**	**1**
	p-value	0.803	0.465	0.910	0.803	0.803	1.000
30	Position	**1**	**6**	**5**	**4**	**3**	**2**
	p-value	1.000	0.383	0.407	0.407	0.407	0.638
	Score	18	25	17	22	26	18

Table G.7: Model confidence sets based on the predictive ability for squared portfolio returns using the loss function defined in (4.5). See the caption of Table G.4 for an explanation of the table entries.

DJIA dataset. Period I. MCS of EPA for the equally weighted portfolio squared returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	Position	**2**	**5**	**4**	6	**1**	**3**
	p-value	0.296	0.296	0.296	0.002	1.000	0.296
10	Position	**3**	**4**	6	**2**	**1**	5
	p-value	0.071	0.071	0.001	0.345	1.000	0.001
15	Position	**1**	6	**2**	5	**3**	**4**
	p-value	1.000	0.005	0.685	0.005	0.326	0.199
20	Position	**2**	6	**1**	5	**3**	**4**
	p-value	0.367	0.006	1.000	0.011	0.367	0.158
25	Position	4	5	**1**	6	**2**	3
	p-value	0.005	0.004	1.000	0.004	0.061	0.049
30	Position	5	6	**2**	**1**	3	4
	p-value	0.002	0.002	0.136	1.000	0.018	0.018
	Score	17	32	16	25	13	23

Table G.8: Model confidence sets based on the predictive ability for squared portfolio returns using the loss function defined in (4.5). See the caption of Table G.4 for an explanation of the table entries.

DJIA dataset. Period I. HIT test of the minimum variance portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	1%	**2.54**	3.07	3.07	2.94	2.67	2.80
	p-value	0.7895	0.9538	**0.9549**	0.9046	0.8707	0.9205
	5%	**6.81***	7.21*	6.94*	6.94*	6.94*	**6.81***
	p-value	0.0170	0.0108	0.0112	0.0112	**0.0223**	0.0172
	10%	11.08**	11.48**	11.75*	11.48*	**10.95***	11.62
	p-value	0.0085	0.0075	0.0211	0.0295	0.0362	**0.0811**
10	1%	**1.60**	1.87	1.74	1.74	**1.60**	**1.60**
	p-value	0.5135	0.7272	**0.9820**	0.5750	0.5126	0.5120
	5%	4.54	**4.67**	4.54	**4.67**	4.54	4.54
	p-value	0.6107	0.5839	**0.7906**	0.6146	0.6113	0.6114
	10%	9.35	8.95	8.68	**9.48**	9.21	9.21
	p-value	0.1459	0.0816	0.1247	0.0738	0.1585	**0.1588**
15	1%	**0.93**	1.07	1.20	1.20	**0.93**	**0.93**
	p-value	0.9965	0.1357	0.0845	0.2323	**0.9968**	0.9968
	5%	4.14	4.41	4.41	**4.54**	4.14	4.14
	p-value	0.2980	0.4356	0.4524	**0.4987**	0.2977	0.2992
	10%	7.61	8.14	8.68	**8.81**	7.88	8.54
	p-value	0.4368	0.6871	**0.8502**	0.5405	0.4536	0.8441
20	1%	**1.20**	1.47	1.47	1.34	1.34	1.47
	p-value	0.2141	0.9301	0.5153	**0.9826**	0.3886	0.5197
	5%	3.47	3.87	**4.14**	4.01	3.47	3.60
	p-value	0.7794	0.7344	0.7765	**0.7980**	0.7794	0.7789
	10%	7.48	8.28	**8.41**	**8.41**	7.61	7.74
	p-value	0.9621	0.7913	0.9138	0.7433	**0.9693**	0.9448
25	1%	**1.34**	1.47	1.87	1.74*	**1.34**	1.60*
	p-value	0.4985	**0.9939**	0.0772	0.0350	0.4985	0.0105
	5%	4.67	4.81	**5.07**	5.21	4.67	4.81
	p-value	0.6049	0.4531	**0.7543**	0.4600	0.6049	0.5051
	10%	9.35	10.68	10.55	**9.88**	9.48	9.75
	p-value	0.7228	0.9699	0.9094	**0.9739**	0.5810	0.7412
30	1%	**1.74**	2.54	2.80	2.40	1.87	2.27
	p-value	0.7757	0.9153	0.9129	0.9175	0.8351	**0.9213**
	5%	**5.74**	6.81	7.34	7.08	6.01	6.54
	p-value	0.9235	0.9682	**0.9890**	0.9019	0.9568	0.6951
	10%	**10.28**	12.28	12.02	11.21	10.68	10.81
	p-value	0.8853	0.6322	**0.9880**	0.7384	0.9599	0.9487

Table G.9: Results of the HIT test. For each assets cardinality n and model we report the average number of VaR violations when this risk measure is computed at the 1, 5, and 10% confidence levels. The percentage printed in red corresponds to the model that yields the closest number to the specified confidence level. The p-values correspond to the F-test on the HIT regression (4.6) with five lags. The highest p-values are marked in bold. When they imply that the null hypothesis of independence is rejected at the 5% (respectively 1%) level, the corresponding average number of VaR violations is marked with * (respectively **).

DJIA dataset. Period I. HIT test of the equally weighted portfolio returns.

n		Scalar	Hadamard	Rank Deficient ($r=1$)	Rank Deficient ($r=2$)	Almon	Almon Shuffle
5	1%	2.27	2.27	2.27	2.27	**2.14**	2.27
	p-value	0.5905	0.6204	**0.6210**	0.6204	0.4791	0.6198
	5%	6.28	6.54	6.54	6.54	**6.01**	6.54
	p-value	**0.7918**	0.6796	0.6793	0.6776	0.5968	0.6796
	10%	**11.88**	**11.88**	**11.88**	**11.88**	**11.88**	12.02
	p-value	**0.3977**	0.2019	0.2019	0.2017	0.3792	0.2615
10	1%	1.47	**1.34**	1.47	**1.34**	1.47	1.47
	p-value	0.9944	0.9936	**0.9944**	0.9931	0.9943	0.9944
	5%	6.28	**6.14**	6.28	**6.14**	6.28	6.28
	p-value	0.0708	**0.1067**	0.0717	0.1066	0.0703	0.0704
	10%	9.48*	**9.61***	9.48*	**9.61***	9.48*	9.35
	p-value	0.0390	0.0333	0.0403	0.0309	0.0382	**0.2340**
15	1%	**1.07**	**1.07**	**1.07**	**1.07**	**1.07**	**1.07**
	p-value	**0.9961**	0.9960	0.9952	0.9952	0.9959	0.9956
	5%	**3.87**	3.60	**3.87**	3.60	**3.87**	**3.87**
	p-value	**0.2759**	0.2613	0.2753	0.2594	0.2758	0.2756
	10%	**6.81**	6.54	6.41	6.41	6.68	6.54
	p-value	**0.1950**	0.1605	0.0975	0.0969	0.1506	0.1652
20	1%	1.20	**0.93***	1.20	**0.93***	1.20	1.20
	p-value	0.3060	0.0453	0.3059	0.0452	0.3061	**0.3066**
	5%	**2.67**	2.54	**2.67**	**2.67**	**2.67**	**2.67**
	p-value	0.3405	0.2177	0.3399	0.3356	**0.3410**	0.3409
	10%	**5.34**	5.21	**5.34**	5.21	**5.34**	**5.34**
	p-value	0.7912	0.7699	0.4568	0.7709	**0.7925**	0.4519
25	1%	**0.80**	0.53	**0.80**	0.53	**0.80**	0.67
	p-value	0.9998	0.9253	0.9998	0.9184	**0.9998**	0.9992
	5%	3.34	3.20	**3.60**	3.07	3.47	3.47
	p-value	0.8904	0.9086	0.7351	**0.9182**	0.8690	0.8687
	10%	6.14	6.01*	**6.28**	6.01*	6.14*	6.14*
	p-value	0.0501	0.0350	**0.0719**	0.0349	0.0499	0.0497
30	1%	0.80	**0.93**	**0.93**	**0.93**	0.80	0.80
	p-value	**0.9954**	0.9759	0.9752	0.9778	0.9952	0.9951
	5%	2.94	2.94	**3.07**	**3.07**	2.94	**3.07**
	p-value	0.4821	0.4812	0.5690	**0.5697**	0.4807	0.5674
	10%	6.01	5.87	6.14	**6.28**	6.14	**6.28**
	p-value	0.6916	**0.9274**	0.7334	0.6628	0.7295	0.6599

Table G.10: Results of the HIT test. For each assets cardinality n and model we report the average number of VaR violations when this risk measure is computed at the 1, 5, and 10% confidence levels. The percentage printed in red corresponds to the model that yields the closest number to the specified confidence level. The p-values correspond to the F-test on the HIT regression (4.6) with five lags. The highest p-values are marked in bold. When they imply that the null hypothesis of independence is rejected at the 5% (respectively 1%) level, the corresponding average number of VaR violations is marked with * (respectively **).

References

[Alm65] S. Almon. The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):178–196, 1965.

[BGO14] Luc Bauwens, Lyudmila Grigoryeva, and Juan-Pablo Ortega. Estimation and empirical performance of non-scalar dynamic conditional correlation models. *In revision for Computational Statistics and Data Analysis*, 2014.

[BR97] R. B. Bapat and T. E. S. Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, 1997.

[CO14] Stéphane Chrétien and Juan-Pablo Ortega. Multivariate GARCH estimation via a Bregman-proximal trust-region method. *Computational Statistics and Data Analysis*, 76:210–236, 2014.

[Eng08] Robert F. Engle. High dimensional dynamic correlations. In J. L. Castle and N. Shephard, editors, *The Methodology and Practice of Econometrics: Papers in Honour of David F Hendry*. Oxford University Press, Oxford, 2008.

[HJ94] Roger A. Horn and Charles R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994.

[LÖ5] Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, 2005.